Supplement: Additional Results and Proofs from Sparse to Dense Functional Data

Álvaro Gajardo[1], Xiongtao Dai[2], and Hans-Georg Müller[1]

[1]*Department of Statistics, University of California, Davis, USA*

[2]*Division of Biostatistics, University of California, Berkeley, USA*

S1. Additional Simulation Results

Here we report simulation results for an additional measure of performance for $\mathcal{P}_{iK}$, where we computed the estimated 2-Wasserstein distance between the empirical distribution of $\hat{F}_{iK}(\beta_0 + \int_{\mathcal{T}} \beta(s)(X_i(s) - \mu(s))ds)$, $i = 1, \ldots, n$, and a uniform distribution on $(0, 1)$. This is of interest, observing that

$F_{1K}(\eta_{1K}), \ldots, F_{nK}(\eta_{nK})$ constitute an i.i.d. sample from a uniform random variable U in $(0, 1)$. A conditioning argument gives $P(F_{iK}(\eta_{iK}) \le p) = E(P(F_{iK}(\eta_{iK}) \le p|\mathbf{X}_i)) = E(P(\eta_{iK} \le F_{iK}^{-1}(p)|\mathbf{X}_i)) = p$, $p \in (0, 1)$. Thus, if we denote by $F_K(\eta_K)$ a generic probability transformation of the linear response

This research was done while Alvaro Gajardo was a PhD student at the University of California, Davis.

η_K one would expect the random variable $F_K(\eta_K)$ to be close to a uniform distribution over $(0, 1)$, in terms of

$$\mathcal{W}_2^2(F_K(\eta_K), U) = \int_0^1 (Q_K(p) - p)^2 dp, \tag{S.1}$$

where Q_K is the quantile function of the random variable $F_K(\eta_K)$. Since the quantities $F_{1K}(\eta_{1K}), \ldots, F_{nK}(\eta_{nK})$ are i.i.d. and share the same distribution with $F_K(\eta_K)$, we may estimate Q_K by the empirical quantile of the $F_{iK}(\eta_{iK})$.

Defining Z_i to be the ith order statistic of the $F_{jK}(\eta_{jK})$, $j = 1, \ldots, n$, a natural estimate $U_{\mathcal{W}}$ of $W_2^2(F_K(\eta_K), U)$ in (S.1) is (Amari and Matsuda, 2021)

$$U_{\mathcal{W}} = \sum_{i=1}^{n} \frac{z_i^2}{n} - z_i \left(\frac{i^2}{n^2} - \frac{(i-1)^2}{n^2} \right) + \frac{1}{3} \left(\frac{i^3}{n^3} - \frac{(i-1)^3}{n^3} \right),$$

and we define $\hat{U}_{\mathcal{W}}$ analogously after replacing population quantities by their estimated versions. The simulation results are in Table S1. One finds that as n increases, the distance $\hat{U}_{\mathcal{W}}$ diminishes, which reflects better performance of the predictive distributions $\mathcal{P}_{iK}$. Higher noise levels lead to worse performance as it becomes harder to estimate population quantities with the same sample size. Similarly, denser designs have a lower average value of $\hat{U}_{\mathcal{W}}$ as expected.

Table S1: Simulation results for the Wasserstein discrepancy against a uniform distribution $\hat{U}_{\mathcal{W}}$ defined through (S.1) for the same settings as in Table 1, displaying the averages of $\hat{U}_{\mathcal{W}}$ based on 2000 simulation runs. Averages are scaled by a factor $1,000$. Smaller discrepancies indicate improved estimation of predictive distributions.

Measurement Error Noise level		Sparsity setting					
Predictor	Response	Very Sparse		Medium Sparse		Dense	
σ	σ_Y	$n = 500$	$n = 2000$	$n = 500$	$n = 2000$	$n = 500$	$n = 2000$
0.5	0.5	1.74	0.62	0.85	0.46	0.76	0.37
	1.0	2.18	0.75	1.22	0.58	1.25	0.52
1.0	0.5	2.95	1.54	1.05	0.44	0.82	0.45

S2. Assumptions and Main Proofs

S2.1 Assumptions

We assume the following regularity conditions (A1)–(A8), which are similar to those in Zhang and Wang (2016) and Dai *et al.* (2018), and are compiled here in one place to facilitate reading. Recall that $w_i = \left(\sum_{j=1}^n n_j\right)^{-1}$ and $v_i = \left(\sum_{j=1}^n n_j(n_j - 1)\right)^{-1}$.

(A1) $K(\cdot)$ is a symmetric probability density function on $[-1, 1]$ and is Lipschitz continuous: There exists $0 < L < \infty$ such that $|K(u) - K(v)| \leq L|u - v|$ for any $u, v \in [0, 1]$.

(A2) $\{T_{ij} : i = 1, \ldots, n,\ j = 1, \ldots, n_i\}$ are i.i.d. copies of a random variable T defined on $\mathcal{T}$, and n_i are regarded as fixed. The density $f(\cdot)$ of T is bounded below and above,

$$0 < m_f \leq \min_{t \in \mathcal{T}} f(t) \leq \max_{t \in \mathcal{T}} f(t) \leq M_f < \infty.$$

Furthermore $f^{(2)}$, the second derivative of $f(\cdot)$, is bounded.

(A3) X, ϵ, and T are independent.

(A4) $\mu^{(2)}(t)$ and $\partial^2\Gamma(s,t)/\partial s^p \partial t^{2-p}$ exist and are bounded on $\mathcal{T}$ and $\mathcal{T} \times \mathcal{T}$, respectively, for $p = 0, \ldots, 2$.

(A5) $h_\mu \to 0$, $\log(n) \sum_{i=1}^n n_i w_i^2 / h_\mu \to 0$ and $\log(n) \sum_{i=1}^n n_i(n_i - 1) w_i^2 \to 0$.

(A6) For some $\alpha > 2$, $E(\sup_{t \in \mathcal{T}} |X(t) - \mu(t)|^\alpha) < \infty$, $E(|\epsilon|^\alpha) < \infty$, and

$$n \left[\sum_{i=1}^n n_i w_i^2 h_\mu + \sum_{i=1}^n n_i(n_i - 1) w_i^2 h_\mu^2 \right] \left[\frac{\log(n)}{n} \right]^{2/\alpha - 1} \to \infty.$$

(A7) $h_G \to 0$, $\log(n) \sum_{i=1}^n n_i(n_i - 1) v_i^2 / h_G^2 \to 0$ and $\log(n) \sum_{i=1}^n n_i(n_i - 1)(n_i - 2) v_i^2 / h_G \to 0$.

(A8) For some $\beta_\gamma > 2$, $E(\sup_{t \in \mathcal{T}} |X(t) - \mu(t)|^{2\beta_\gamma}) < \infty$, $E(|\epsilon|^{2\beta_\gamma}) < \infty$, and

$$n \Bigg[\sum_{i=1}^n n_i(n_i - 1) v_i^2 h_G^2 + \sum_{i=1}^n n_i(n_i - 1)(n_i - 2) v_i^2 h_G^3 + \sum_{i=1}^n n_i(n_i - 1)(n_i - 2)(n_i - 3) v_i^2 h_G^4 \Bigg] \left[\frac{\log(n)}{n} \right]^{2/\beta_\gamma - 1} \to \infty.$$

We remark that assumption (A2) implies (X1) in the main text and assumption (A4) implies (X3).

S2.2 Additional Details for Mean and Covariance Estimation

For notational simplicity, for a function $g_1 : \mathcal{T} \to \mathbb{R}$ and a vector $\boldsymbol{z} = (z_1, \dots, z_p)^T \in \mathbb{R}^p$, $p > 0$, denote by $g_1(\boldsymbol{z}) = (g_1(z_1), \dots, g_1(z_p))^T$ the application of g_1 to $\boldsymbol{z}$ entry-wise. Similarly, for a function $g_2 : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ and a second vector $\boldsymbol{r} = (r_1, \dots, r_q)^T \in \mathbb{R}^q$, $q > 0$, denote by $g_2(\boldsymbol{z}, \boldsymbol{r}^T)$ the $p \times q$ matrix, for which the (l, k) element is given by $g_2(z_l, r_k)$, where $1 \le l \le p$ and $1 \le k \le q$. Also, for two scalar sequences θ_n and γ_n, write $\theta_n \lesssim \gamma_n$ if there exists a constant $c_0 > 0$ such that $\theta_n \le c_0\gamma_n$ holds for large enough n.

For the mean function estimate, set $\hat{\mu}(t) = \hat{\gamma}_0$, where

$$(\hat{\gamma}_0, \hat{\gamma} = \underset{\gamma_0,\gamma_1}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \sum_{j=1}^{n_i} (X_{ij} - \gamma_0 - \gamma_1(T_{ij} - t))^2 K_{h_\mu}(T_{ij} - t),$$

where $w_i = (\sum_{j=1}^{n} n_j)^{-1}$ are equal subject weights, K is a kernel function corresponding to a density function with compact support $[-1, 1]$ and $K_{h_\mu}(\cdot) = K(\cdot/h_\mu)/h_\mu$. For the covariance surface estimate, writing $\hat{C}_{ijl} = (X_{ij} - \hat{\mu}(T_{ij}))(X_{il} - \hat{\mu}(T_{il}))$ for the raw covariances (Yao *et al.*, 2005), set $\hat{\Gamma}(s, t) = \hat{\gamma}_0$, where

$$\begin{aligned}
&(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2) \\
&= \underset{\gamma_0,\gamma_1,\gamma_2}{\operatorname{argmin}} \sum_{i=1}^{n} v_i \sum_{1 \le j \ne l \le n_i} (C_{ijl} - \gamma_0 - \gamma_1(T_{ij} - s) - \gamma_2(T_{il} - t))^2 \\
&\times K_{h_G}(T_{ij} - s) K_{h_G}(T_{il} - t).
\end{aligned}$$

Here $v_i = (\sum_{j=1}^{n} n_j(n_j - 1))^{-1}$ and $n_i \ge 2$ is assumed throughout for the covariance estimation step.

For the cross-covariance smoothing step, using the raw covariances $C_i(T_{ij}) =$

$(\tilde{X}_{ij} - \hat{\mu}(T_{ij}))Y_i$, the local linear estimate of $C(t)$ is given by $\hat{C}(t) = \hat{\beta}_0^X$, where

$$(\hat{\beta}_0^X, \hat{\beta}_1^X) = \underset{\beta_0^X, \beta_1^X \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{n_i} w_i K_h(T_{ij} - t)(C_i(T_{ij}) - \beta_0^X - \beta_1^X (t - T_{ij}))^2, \tag{S.2}$$

with $w_i = (\sum_{i=1}^{n} n_i)^{-1}$.

S2.3 Proofs of Main Results in Section 2 and 3

The proofs in this and the following sections rely on various auxiliary results and lemmas included in Section S3.

Proof of Proposition 1. Fix $i \in \{1, \dots, n\}$ and $k \in \mathbb{N}$, and recall that

$$\tilde{\xi}_{ik} = \lambda_k \boldsymbol{\phi}_{ik}^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{X}_i - \boldsymbol{\mu}_i), \tag{S.3}$$

where $\boldsymbol{\phi}_{ik} = (\phi_k(T_{i1}), \dots, \phi_k(T_{im}))^T$. Define $\mathbf{W} = \operatorname{diag}(w_l)$, where w_l are quadrature weights chosen according to the left endpoint rule, i.e. $w_l = T_{il} - \max_{j:T_{ij}<T_{il}} T_{ij}$ for $l = 1, \dots, m$, and we set $\max_{j:T_{ij}<T_{il}} T_{ij} = 0$ whenever $\{j : T_{ij} < T_{il}\} = \varnothing$. Let g_m be the size of the maximal gap between $\{0, T_{i1}, \dots, T_{im}, 1\}$ for $\mathcal{T} = [0, 1]$ and consider the quadrature approximation errors

$$\mathbf{e}_k = \int_{\mathcal{T}} \Gamma(\mathbf{T}_i, t)\phi_k(t)dt - \boldsymbol{\Sigma}_i \mathbf{W} \boldsymbol{\phi}_{ik},$$

where $\Gamma(\mathbf{T}_i, t) = (\Gamma(T_{i1}, t), \dots, \Gamma(T_{im}, t))^T$. Here note that since $\boldsymbol{\Sigma}_i = \sigma^2 I_m + \Gamma(\mathbf{T}_i, \mathbf{T}_i^T)$, where $\Gamma(\mathbf{T}_i, \mathbf{T}_i^T)$ corresponds to the matrix with elements $[\Gamma(\mathbf{T}_i, \mathbf{T}_i^T)]_{jl} = \Gamma(T_{ij}, T_{il})$, $j, l \in \{1, \dots, m\}$, we have $\boldsymbol{\Sigma}_i \mathbf{W} \boldsymbol{\phi}_{ik} = \sigma^2 \mathbf{W} \boldsymbol{\phi}_{ik} + \Gamma(\mathbf{T}_i, \mathbf{T}_i^T) \mathbf{W} \boldsymbol{\phi}_{ik}$ where the second term in the previous expression corresponds to the numerical quadrature approximation to $\int_{\mathcal{T}} \Gamma(\mathbf{T}_i, t)\phi_k(t)dt$ and the first term will be shown to be negligible as $m \to \infty$.

From the quadrature approximation error for integrating a continuously differentiable function g over $[0, 1]$ under the left-endpoint rule and denoting $T_i^{(m)} := \max_{1\leq j\leq m} T_{ij}$ we have

$$\left|\int_0^1 g(t)dt - \sum_{l=1}^m g(T_{il})w_l\right|$$
$$\leq \frac{\sup_{t\in\mathcal{T}} |g'(t)|}{2}\left(\sum_{l=1}^m w_l^2 + (1 - T_i^{(m)})^2\right) + |(1 - T_i^{(m)})g(1)| \tag{S.4}$$
$$= O_p(m^{-1}), \tag{S.5}$$

where (S.5) follows from Lemma S2. Denoting by $\|\cdot\|_2$ the Euclidean norm in $\mathbb{R}^m$, we have

$$\|\mathbf{e}_k\|_2 \leq \left\|\int_{\mathcal{T}} \Gamma(\mathbf{T}_i, t)\phi_k(t)dt - \Gamma(\mathbf{T}_i, \mathbf{T}_i^T)\mathbf{W}\boldsymbol{\phi}_{ik}\right\|_2 + \left\|\sigma^2\mathbf{W}\boldsymbol{\phi}_{ik}\right\|_2 = O_p(m^{-1/2}), \tag{S.6}$$

which follows by noting that the integration error rates for all entries in $\mathbf{e}_k$ are uniform due to Condition (X3) in the main text and (S.4), and that

$$\|\mathbf{W}\boldsymbol{\phi}_{ik}\|_2^2 \leq \sum_{l=1}^m w_l^2 \sup_{t\in\mathcal{T}} \phi_k^2(t) = O_p(m^{-1}). \tag{S.7}$$

Since

$$\lambda_k\boldsymbol{\phi}_{ik} = \boldsymbol{\Sigma}_i\mathbf{W}\boldsymbol{\phi}_{ik} + \mathbf{e}_k, \tag{S.8}$$

we have

$$\lambda_k \boldsymbol{\phi}_{ik}^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{X}_i - \boldsymbol{\mu}_i) = \boldsymbol{\phi}_{ik}^T \mathbf{W}(\mathbf{X}_i - \boldsymbol{\mu}_i) + \mathbf{e}_k^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{X}_i - \boldsymbol{\mu}_i)$$
$$= \boldsymbol{\phi}_{ik}^T \mathbf{W}(\mathbf{Y}_i - \boldsymbol{\mu}_i) + \boldsymbol{\phi}_{ik}^T \mathbf{W}\boldsymbol{\epsilon}_i + \mathbf{e}_k^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{X}_i - \boldsymbol{\mu}_i), \tag{S.9}$$

where $\mathbf{Y}_i = (X_{i1}, \ldots, X_{im})^T$ and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{im})^T$. Let $g_k(t) = \phi_k(t)(X_i(t) - \mu(t))$. Then, from Condition (X3) and since the process $X_i(t)$ is assumed continuously differentiable almost surely, we have $g_k(t)$ is continuously differentiable a.s. over the compact set $\mathcal{T} = [0, 1]$ so that $\sup_{t\in\mathcal{T}} |g_k'(t)| = O_p(1)$. Thus, using (S.4) and the fact that $\int_0^1 \phi_k(t)(X_i(t) - \mu(t))dt = \xi_{ik}$, we obtain

$$\xi_{ik} - \sum_{l=1}^{m} \phi_k(T_{il})(X_i(T_{il}) - \mu(T_{il}))w_l = O_p(m^{-1}),$$

whence

$$\boldsymbol{\phi}_{ik}^T \mathbf{W}(\mathbf{Y}_i - \boldsymbol{\mu}_i) = \xi_{ik} + O_p(m^{-1}). \tag{S.10}$$

By conditioning and using the independence between $\boldsymbol{\epsilon}_i$ and $\mathbf{T}_i$, $E(\boldsymbol{\phi}_{ik}^T \mathbf{W}\boldsymbol{\epsilon}_i)^2 = E[E(\boldsymbol{\phi}_{ik}^T \mathbf{W}\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^T \mathbf{W}\boldsymbol{\phi}_{ik}|\mathbf{T}_i)]$
$= E[\boldsymbol{\phi}_{ik}^T \mathbf{W} E(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^T)\mathbf{W}\boldsymbol{\phi}_{ik}] = \sigma^2 E(\|\mathbf{W}\boldsymbol{\phi}_{ik}\|_2^2)$.

Hence, from (S.7) it follows that $E(\boldsymbol{\phi}_{ik}^T \mathbf{W}\boldsymbol{\epsilon}_i)^2 = O(m^{-1})$ and thus

$$\boldsymbol{\phi}_{ik}^T \mathbf{W}\boldsymbol{\epsilon}_i = O_p(m^{-1/2}). \tag{S.11}$$

We now show that $Z_m := \mathbf{e}_k^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{X}_i - \boldsymbol{\mu}_i) = O_p(m^{-1/2})$. Note that for any $M > 0$

$$P\left(\sqrt{m}\,|Z_m| > M|\mathbf{T}_i\right) \le \frac{m}{M^2}\|\mathbf{e}_k\|_2^2 \|\boldsymbol{\Sigma}_i^{-1/2}\|_{\text{op},2}^2 \le \frac{m}{M^2\sigma^2}\|\mathbf{e}_k\|_2^2, \tag{S.12}$$

where the last inequality follows since $\|\mathbf{\Sigma}_i^{-1/2}\|_{\text{op},2} \leq \sigma^{-1}$. From (S.6), $m\,\|\mathbf{e}_k\|_2^2 = O_p(1)$ and thus for any $\epsilon > 0$ there exist $M_0 = M_0(\epsilon) > 0$ and $m_0 = m_0(\epsilon) \in \mathbb{N}^+$ such that

$$P\left(m\,\|\mathbf{e}_k\|_2^2 > M_0\right) \leq \epsilon, \quad \forall m \geq m_0. \tag{S.13}$$

Hence, by choosing $M = M_\epsilon := \sqrt{M_0/(\epsilon\sigma^2)}$ and defining $u_{im} := P\left(\sqrt{m}\,|Z_m| > M|\mathbf{T}_i\right)$,

$$\begin{aligned} P\left(\sqrt{m}\,|Z_m| > M_\epsilon\right) &= E[u_{im}] \\ &= E[u_{im}1_{\{u_{im}\leq\epsilon\}} + u_{im}1_{\{u_{im}>\epsilon\}}] \leq \epsilon + P(u_{im} > \epsilon), \end{aligned} \tag{S.14}$$

where the last inequality follows since $u_{im} \leq 1$. Now (S.12) and (S.13) imply $P(u_{im} > \epsilon) \leq \epsilon$ for $m \geq m_0$, whence

$$P\left(\sqrt{m}\,|\mathbf{e}_k^T\mathbf{\Sigma}_i^{-1}(\mathbf{X}_i - \boldsymbol{\mu}_i)| > M_\epsilon\right) \leq 2\epsilon, \quad \forall m \geq m_0, \tag{S.15}$$

which shows that $\mathbf{e}_k^T\mathbf{\Sigma}_i^{-1}(\mathbf{X}_i - \boldsymbol{\mu}_i) = O_p(m^{-1/2})$. The result follows by combining (S.9), (S.10), (S.11) and (S.15). □

Proof of Theorem 1. Let $K_0 \geq k$ be any fixed integer and consider the constant sequence $K = K(n) = K_0$, for all $n \geq 1$. Thus $(a_n + b_n)\sum_{k=1}^{K}\lambda_k^{-1} = o(1)$ as $n \to \infty$ and similar arguments as the ones outlined in the proof of Lemma S3 leads to

$$\begin{aligned} (\hat{\xi}_k^* - \tilde{\xi}_k^*)^2 &\lesssim (\hat{\mathbf{e}}_k^{*T}\hat{\mathbf{\Sigma}}^{*-1}(\mathbf{X}^* - \hat{\boldsymbol{\mu}}^*))^2 + (\hat{\boldsymbol{\phi}}_k^{*T}\mathbf{W}^*(\mathbf{X}^* - \hat{\boldsymbol{\mu}}^*) - \boldsymbol{\phi}_k^{*T}\mathbf{W}^*(\mathbf{X}^* - \boldsymbol{\mu}^*))^2 \\ &\quad + (\mathbf{e}_k^{*T}\mathbf{\Sigma}^{*-1}(\mathbf{X}^* - \boldsymbol{\mu}^*))^2 \\ &= O_p\left(m^{*-1} + m^{*2}(a_n + b_n)^2\right), \end{aligned}$$

where $\hat{\mathbf{e}}_k^*$ is defined as in (S.75). The result follows. □

Proof of Theorem 3. Recall that $\tilde{\mu}_{iK} = \tilde{\boldsymbol{\xi}}_{iK}^T \boldsymbol{\Phi}_K$ and $K = K(m)$ satisfies $\sum_{k=1}^K \lambda_k^{-1} \asymp m^{1-\delta}$, where $\delta \in (1/2, 1)$ and $\tilde{\boldsymbol{\xi}}_{iK} = \boldsymbol{\Lambda}_K \boldsymbol{\Phi}_{iK}^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{X}_i - \boldsymbol{\mu}_i)$. We first show shrinkage of $\|\tilde{\mu}_{iK} - \sum_{k=1}^\infty \xi_{ik}\phi_k\|_{L^2}$. Also, for any $k \geq 1$ define

$$\mathbf{e}_k = \int_{\mathcal{T}} \Gamma(\mathbf{T}_i, t)\phi_k(t)dt - \boldsymbol{\Sigma}_i \mathbf{W} \boldsymbol{\phi}_{ik}.$$

From (S.9) and the triangle inequality, we have

$$\begin{aligned}
\|\tilde{\mu}_{iK} - \sum_{k=1}^\infty \xi_{ik}\phi_k\|_{L^2} &= \|\sum_{k=1}^K \lambda_k \boldsymbol{\phi}_{ik}^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{X}_i - \boldsymbol{\mu}_i)\phi_k - \sum_{k=1}^\infty \xi_{ik}\phi_k\|_{L^2} \\
&\leq \|\sum_{k=1}^K \boldsymbol{\phi}_{ik}^T \mathbf{W}(\mathbf{Y}_i - \boldsymbol{\mu}_i)\phi_k - \sum_{k=1}^\infty \xi_{ik}\phi_k\|_{L^2} + \|\sum_{k=1}^K \boldsymbol{\phi}_{ik}^T \mathbf{W}\boldsymbol{\epsilon}_i \phi_k\|_{L^2} \\
&+ \|\sum_{k=1}^K \mathbf{e}_k^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{X}_i - \boldsymbol{\mu}_i)\phi_k\|_{L^2} \\
&= \|A\|_{L^2} + \|B\|_{L^2} + \|C\|_{L^2}, \qquad \text{(S.16)}
\end{aligned}$$

where the functions $A = A(t)$, $B = B(t)$ and $C = C(t)$ are defined through the last equation. By Fubini's theorem and orthogonality of the ϕ_k, we have

$$\begin{aligned}
&E(\|B\|_{L^2}^2) \\
&= \int_{\mathcal{T}} E\left[\left(\sum_{k=1}^K \phi_k(t)\boldsymbol{\phi}_{ik}^T \mathbf{W}\boldsymbol{\epsilon}_i\right)^2\right] dt = \sum_{k=1}^K E[(\boldsymbol{\phi}_{ik}^T \mathbf{W}\boldsymbol{\epsilon}_i)^2] = \sum_{k=1}^K \sigma^2 E(\|\mathbf{W}\boldsymbol{\phi}_{ik}\|_2^2),
\end{aligned}$$

where the last equality follows from the proof of Theorem 1. Thus, from (S.7) and Lemma S2 we obtain

$$\begin{aligned}
&E(\|B\|_{L^2}^2) \\
&\leq \sum_{k=1}^K \sigma^2 m^{-1}\|\phi_k\|_\infty^2 = O\left(m^{-1}\sum_{k=1}^K \lambda_k^{-2}\right) = O\left(m^{-1}\Big[\sum_{k=1}^K \lambda_k^{-1}\Big]^2\right) = O(m^{1-2\delta}),
\end{aligned}$$

where the first equality is due to $\|\phi_k\|_\infty = O(\lambda_k^{-1})$. This follows from the relation

$$\lambda_k \phi_k(t) = \int_{\mathcal{T}} \Gamma(t,s)\phi_k(s)ds \leq \|\Gamma(t,\cdot)\|_{L^2} < \infty$$

uniformly over t, which is a consequence of the Cauchy–Schwarz inequality and continuity of Γ over the compact set $\mathcal{T}^2$. Therefore

$$\|B\|_{L^2} = O_p(m^{1/2-\delta}). \tag{S.17}$$

Observe

$$A(t) = \sum_{k=1}^{K} (\boldsymbol{\phi}_{ik}^T \mathbf{W}(\mathbf{Y}_i - \boldsymbol{\mu}_i) - \xi_{ik})\phi_k(t) - \sum_{k=K+1}^{\infty} \xi_{ik}\phi_k(t) = A_1(t) - A_2(t),$$

where $A_1(t)$ and $A_2(t)$ are defined through the last equation. By Fubini's theorem along with the orthonormality of the ϕ_k, we have

$$E(\|A_2\|_{L^2}^2) = \sum_{k=K+1}^{\infty} \lambda_k,$$

and then

$$\|A_2\|_{L^2} = O_p\left(\left(\sum_{k=K+1}^{\infty} \lambda_k\right)^{1/2}\right). \tag{S.18}$$

Define $g_k(t) = \phi_k(t)(X_i(t) - \mu(t))$, $t \in \mathcal{T}$. By the dominated convergence theorem along with the Cauchy–Schwarz inequality,

$$\lambda_k |\phi_k'(t)| = \left| \int_{\mathcal{T}} \Gamma^{(1,0)}(t,s)\phi_k(s)ds \right| \leq \|\Gamma^{(1,0)}\|_\infty < \infty,$$

where $\Gamma^{(1,0)}(t,s) = \partial\Gamma(t,s)/\partial t$. This shows that $\|\phi_k'\|_\infty = O(\lambda_k^{-1})$ which combined with the fact that $\|\phi_k\|_\infty = O(\lambda_k^{-1})$ and Condition (X2) leads to

$\|g_k'\|_\infty = O(\lambda_k^{-1})$ and $\|g_k\|_\infty = O(\lambda_k^{-1})$. Hence, from the Riemann sum approximation error bound in (S.4) applied to the function $g_k(t) = \phi_k(t)(X_i(t) - \mu(t))$, we obtain

$$|\boldsymbol{\phi}_{ik}^T \mathbf{W}(\mathbf{Y}_i - \boldsymbol{\mu}_i) - \xi_{ik}| \lesssim \lambda_k^{-1} \left(\sum_{l=1}^m w_l^2 + (1 - T_i^{(m)})^2 + (1 - T_i^{(m)}) \right).$$

Therefore

$$E(\|A_1\|_{L^2}) \leq \sum_{k=1}^K E(|\boldsymbol{\phi}_{ik}^T \mathbf{W}(\mathbf{Y}_i - \boldsymbol{\mu}_i) - \xi_{ik}|) \lesssim \sum_{k=1}^K \lambda_k^{-1} m^{-1} = O(m^{-\delta}),$$

where we use the condition $\sum_{k=1}^K \lambda_k^{-1} \asymp m^{1-\delta}$. This shows that $\|A_1\|_{L^2} = O_p(m^{-\delta})$, which combined with (S.18) leads to

$$\|A\|_{L^2} = O_p\left(m^{-\delta} + \left(\sum_{k=K+1}^\infty \lambda_k \right)^{1/2} \right). \tag{S.19}$$

From (S.6), (S.7), the Riemann sum approximation error bound (S.4), and using that $\|\phi_k'\|_\infty = O(\lambda_k^{-1})$ along with $\|\phi_k\|_\infty = O(\lambda_k^{-1})$, we obtain

$$\|\mathbf{e}_k\|_2 \lesssim \sqrt{m}\lambda_k^{-1} \left(\sum_{l=1}^m w_l^2 + (1 - T_i^{(m)})^2 + (1 - T_i^{(m)}) \right) + \lambda_k^{-1} \left(\sum_{l=1}^m w_l^2 \right)^{1/2}, \tag{S.20}$$

Thus, using the inequality $(x_0 + x_1)^2 \leq 2x_0^2 + 2x_1^2$, which is valid for all $x_0, x_1 \in \mathbb{R}$, along with Lemma S2 leads to

$$\begin{aligned} E(\|\mathbf{e}_k\|_2^2) &\lesssim E\left(m\lambda_k^{-2} \left(\left(\sum_{l=1}^m w_l^2 \right)^2 + (1 - T_i^{(m)})^4 + (1 - T_i^{(m)})^2 \right) + \lambda_k^{-2} \sum_{l=1}^m w_l^2 \right) \\ &= O(m^{-1}\lambda_k^{-2}). \end{aligned} \tag{S.21}$$

Therefore

$$E(\|C\|_{L^2}) \leq \sum_{k=1}^{K} E(|\mathbf{e}_k^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{X}_i - \boldsymbol{\mu}_i)|) \leq \sum_{k=1}^{K} (E\{E[(\mathbf{e}_k^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{X}_i - \boldsymbol{\mu}_i))^2 | \mathbf{T}_i]\})^{1/2}$$
$$\leq \sigma^{-1} \sum_{k=1}^{K} (E(\|\mathbf{e}_k\|_2^2))^{1/2} \lesssim m^{1/2-\delta},$$

where last inequality uses that $\sum_{k=1}^{K} \lambda_k^{-1} \asymp m^{1-\delta}$. Hence

$$\|C\|_{L^2} = O_p(m^{1/2-\delta}). \tag{S.22}$$

Combining (S.16), (S.17), (S.19), and (S.22) leads to

$$\|\tilde{\mu}_{iK} - \sum_{k=1}^{\infty} \xi_{ik}\phi_k\|_{L^2} = O_p\left(m^{1/2-\delta} + \left(\sum_{k=K+1}^{\infty} \lambda_k\right)^{1/2}\right). \tag{S.23}$$

By orthonormality of the ϕ_k and since $\boldsymbol{\Sigma}_{iK} = \boldsymbol{\Lambda}_K - \boldsymbol{\Lambda}_K \boldsymbol{\Phi}_{iK}^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\Phi}_{iK} \boldsymbol{\Lambda}_K$,

$$\int_{\mathcal{T}} \Gamma_{iK}(t,t)dt = \text{trace}(\boldsymbol{\Sigma}_{iK}) = \sum_{k=1}^{K} \left(\lambda_k - \lambda_k \boldsymbol{\phi}_{ik}^T \boldsymbol{\Sigma}_i^{-1} \lambda_k \boldsymbol{\phi}_{ik}\right). \tag{S.24}$$

From (S.21) and using the condition $\sum_{k=1}^{K} \lambda_k^{-1} \asymp m^{1-\delta}$, we obtain $\sum_{k=1}^{K} \lambda_k^{-2} = O(m^{2-2\delta})$ and

$$E\left(\sum_{k=1}^{K} \mathbf{e}_k^T \boldsymbol{\Sigma}_i^{-1} \mathbf{e}_k\right) \leq \sigma^{-2} \sum_{k=1}^{K} E(\|\mathbf{e}_k\|_2^2) = O(m^{1-2\delta}).$$

Thus

$$\sum_{k=1}^{K} \mathbf{e}_k^T \boldsymbol{\Sigma}_i^{-1} \mathbf{e}_k = O_p(m^{1-2\delta}). \tag{S.25}$$

Since $\|\phi_k\|_\infty = O(\lambda_k^{-1})$ and $\sum_{k=1}^{K} \lambda_k^{-2} = O(m^{2-2\delta})$,

$$\sum_{k=1}^{K} \lambda_k^{-1} \|\mathbf{e}_k\|_2 \lesssim m^{5/2-2\delta} \left(\sum_{l=1}^{m} w_l^2 + (1 - T_i^{(m)})^2 + (1 - T_i^{(m)})\right)$$
$$+ m^{2-2\delta} \left(\sum_{l=1}^{m} w_l^2\right)^{1/2} = O_p\left(m^{3/2-2\delta}\right),$$

where the first inequality is due to (S.20) and the last equality is due to Lemma S2. Thus

$$\sum_{k=1}^{K}|\mathbf{e}_k^T\mathbf{W}\boldsymbol{\phi}_{ik}| \le \sum_{k=1}^{K}\|\mathbf{e}_k\|_2\|\mathbf{W}\boldsymbol{\phi}_{ik}\|_2 \le \left(\sum_{l=1}^{m} w_l^2\right)^{1/2}\sum_{k=1}^{K}\|\mathbf{e}_k\|_2\|\phi_k\|_\infty = O_p\left(m^{1-2\delta}\right), \tag{S.26}$$

where the second inequality is due to (S.7). Also,

$$\sum_{k=1}^{K}\sigma^2|\boldsymbol{\phi}_{ik}^T\mathbf{W}\mathbf{W}\boldsymbol{\phi}_{ik}| \le \sigma^2\sum_{k=1}^{K}\|\mathbf{W}\boldsymbol{\phi}_{ik}\|_2^2 \le \sigma^2\sum_{k=1}^{K}\|\phi_k\|_\infty^2\left(\sum_{l=1}^{m} w_l^2\right) = O_p\left(m^{1-2\delta}\right). \tag{S.27}$$

From the Riemann sum approximation error bound (S.4) applied to the function $g_k(t) = \lambda_k\phi_k^2(t)$, and using that $\|g_k\|_\infty = O(\lambda_k^{-1})$ and $\|g_k'\|_\infty = O(\lambda_k^{-1})$, we have

$$|\lambda_k\boldsymbol{\phi}_{ik}^T\mathbf{W}\boldsymbol{\phi}_{ik} - \lambda_k| = O\left(\lambda_k^{-1}\left(\sum_{l=1}^{m} w_l^2 + (1-T_i^{(m)})^2 + (1-T_i^{(m)})\right)\right).$$

Thus

$$E\left(\sum_{k=1}^{K}|\lambda_k\boldsymbol{\phi}_{ik}^T\mathbf{W}\boldsymbol{\phi}_{ik} - \lambda_k|\right) = O(m^{-\delta}),$$

which implies

$$\sum_{k=1}^{K}|\lambda_k\boldsymbol{\phi}_{ik}^T\mathbf{W}\boldsymbol{\phi}_{ik} - \lambda_k| = O_p\left(m^{-\delta}\right). \tag{S.28}$$

Also, from (S.4) and (S.7) we have

$$\begin{aligned}
&\sum_{k=1}^{K}|\boldsymbol{\phi}_{ik}^T\mathbf{W}\left(\Gamma(\mathbf{T}_i,\mathbf{T}_i^T)\mathbf{W}\boldsymbol{\phi}_{ik} - \lambda_k\boldsymbol{\phi}_{ik}\right)| \\
&\le \sum_{k=1}^{K}\|\boldsymbol{\phi}_{ik}^T\mathbf{W}\|_2\|\Gamma(\mathbf{T}_i,\mathbf{T}_i^T)\mathbf{W}\boldsymbol{\phi}_{ik} - \lambda_k\boldsymbol{\phi}_{ik}\|_2 \\
&\lesssim \sum_{k=1}^{K}\lambda_k^{-2}m^{1/2}\left(\sum_{l=1}^{m} w_l^2\right)^{1/2}\left(\sum_{l=1}^{m} w_l^2 + (1-T_i^{(m)})^2 + (1-T_i^{(m)})\right),
\end{aligned}$$

which along with Lemma S2 leads to

$$E\left(\sum_{k=1}^{K}|\boldsymbol{\phi}_{ik}^T\mathbf{W}\left(\Gamma(\mathbf{T}_i,\mathbf{T}_i^T)\mathbf{W}\boldsymbol{\phi}_{ik}-\lambda_k\boldsymbol{\phi}_{ik}\right)|\right)=O(m^{1-2\delta}).$$

This shows that

$$\sum_{k=1}^{K}[\boldsymbol{\phi}_{ik}^T\mathbf{W}\left(\Gamma(\mathbf{T}_i,\mathbf{T}_i^T)\mathbf{W}\boldsymbol{\phi}_{ik}-\lambda_k\boldsymbol{\phi}_{ik}\right)]=O_p(m^{1-2\delta}). \tag{S.29}$$

From (S.25), (S.26), (S.27), (S.28), (S.29), and observing

$$\boldsymbol{\phi}_{ik}^T\mathbf{W}\boldsymbol{\Sigma}_i\mathbf{W}\boldsymbol{\phi}_{ik}=\sigma^2\boldsymbol{\phi}_{ik}^T\mathbf{W}\mathbf{W}\boldsymbol{\phi}_{ik}+\boldsymbol{\phi}_{ik}^T\mathbf{W}\Gamma(\mathbf{T}_i,\mathbf{T}_i^T)\mathbf{W}\boldsymbol{\phi}_{ik},$$

leads to

$$\begin{aligned}
&\Big|\sum_{k=1}^{K}(\lambda_k-\lambda_k\boldsymbol{\phi}_{ik}^T\boldsymbol{\Sigma}_i^{-1}\lambda_k\boldsymbol{\phi}_{ik})\Big|\\
&=\Big|\sum_{k=1}^{K}(\lambda_k-\mathbf{e}_k^T\boldsymbol{\Sigma}_i^{-1}\mathbf{e}_k-2\mathbf{e}_k^T\mathbf{W}\boldsymbol{\phi}_{ik}-\boldsymbol{\phi}_{ik}^T\mathbf{W}\boldsymbol{\Sigma}_i\mathbf{W}\boldsymbol{\phi}_{ik})\Big|\\
&\le\sum_{k=1}^{K}\mathbf{e}_k^T\boldsymbol{\Sigma}_i^{-1}\mathbf{e}_k+2\sum_{k=1}^{K}|\mathbf{e}_k^T\mathbf{W}\boldsymbol{\phi}_{ik}|+\sigma^2\sum_{k=1}^{K}|\boldsymbol{\phi}_{ik}^T\mathbf{W}\mathbf{W}\boldsymbol{\phi}_{ik}|\\
&+\Big|\sum_{k=1}^{K}[\boldsymbol{\phi}_{ik}^T\mathbf{W}\left(\Gamma(\mathbf{T}_i,\mathbf{T}_i^T)\mathbf{W}\boldsymbol{\phi}_{ik}-\lambda_k\boldsymbol{\phi}_{ik}\right)]\Big|+\sum_{k=1}^{K}|\lambda_k\boldsymbol{\phi}_{ik}^T\mathbf{W}\boldsymbol{\phi}_{ik}-\lambda_k|\\
&=O_p\left(m^{1-2\delta}\right),
\end{aligned}$$

where the first equality uses (S.8). This along with (S.24) implies

$$\int_{\mathcal{T}}\Gamma_{iK}(t,t)dt=O_p\left(m^{1-2\delta}\right). \tag{S.30}$$

Combining (S.30) with (S.23) leads to the result. □

S3. Auxiliary Results and Proofs

We provide the proofs of Propositions 2 and Theorems 2, 4–7 in the main text, followed by a sequence of auxiliary lemmas and their proofs. These auxiliary results are used to derive the main results.

Proof of Proposition 2. Recalling that $\mathbf{\Sigma}_{iK} = \mathbf{\Lambda}_K - \mathbf{\Lambda}_K \mathbf{\Phi}_{iK}^T \mathbf{\Sigma}_i^{-1} \mathbf{\Phi}_{iK} \mathbf{\Lambda}_K$ we have

$$\|\mathbf{\Sigma}_{iK}\|_{\text{op},2} \leq \text{trace}(\mathbf{\Sigma}_{iK}) = \sum_{k=1}^{K} \left(\lambda_k - \lambda_k \boldsymbol{\phi}_{ik}^T \mathbf{\Sigma}_i^{-1} \lambda_k \boldsymbol{\phi}_{ik}\right). \tag{S.31}$$

Moreover, since $\lambda_k \boldsymbol{\phi}_{ik} = \mathbf{e}_k + \mathbf{\Sigma}_i \mathbf{W} \boldsymbol{\phi}_{ik}$, where $\mathbf{e}_k$ is defined as in the proof of Proposition 1, it follows that

$$\lambda_k \boldsymbol{\phi}_{ik}^T \mathbf{\Sigma}_i^{-1} \lambda_k \boldsymbol{\phi}_{ik} = \mathbf{e}_k^T \mathbf{\Sigma}_i^{-1} \mathbf{e}_k + 2\mathbf{e}_k^T \mathbf{W} \boldsymbol{\phi}_{ik} + \boldsymbol{\phi}_{ik}^T \mathbf{W} \mathbf{\Sigma}_i \mathbf{W} \boldsymbol{\phi}_{ik}. \tag{S.32}$$

From, (S.6),

$$\left\|\lambda_k \boldsymbol{\phi}_{ik} - \Gamma(\mathbf{T}_i, \mathbf{T}_i^T) \mathbf{W} \boldsymbol{\phi}_{ik}\right\|_2 = O_p(m^{-1/2}),$$

and using (S.7),

$$\begin{aligned} \boldsymbol{\phi}_{ik}^T \mathbf{W} \mathbf{\Sigma}_i \mathbf{W} \boldsymbol{\phi}_{ik} &= \sigma^2 \boldsymbol{\phi}_{ik}^T \mathbf{W} \mathbf{W} \boldsymbol{\phi}_{ik} + \boldsymbol{\phi}_{ik}^T \mathbf{W} \Gamma(\mathbf{T}_i, \mathbf{T}_i^T) \mathbf{W} \boldsymbol{\phi}_{ik} \\ = O_p(m^{-1}) + \boldsymbol{\phi}_{ik}^T \mathbf{W} &\left(\lambda_k \boldsymbol{\phi}_{ik} - O_p(m^{-1/2})\right) = \lambda_k \boldsymbol{\phi}_{ik}^T \mathbf{W} \boldsymbol{\phi}_{ik} + O_p(m^{-1}), \end{aligned}$$

where $\lambda_k \boldsymbol{\phi}_{ik}^T \mathbf{W} \boldsymbol{\phi}_{ik} = \lambda_k + O_p(m^{-1})$. This follows from the quadrature approximation error (S.5), observing $\int_0^1 \phi_k^2(t) dt = 1$, and implies

$$\boldsymbol{\phi}_{ik}^T \mathbf{W} \mathbf{\Sigma}_i \mathbf{W} \boldsymbol{\phi}_{ik} = \lambda_k + O_p(m^{-1}). \tag{S.33}$$

The result then follows by combining (S.31), (S.32), (S.33), (S.6), (S.7), and the fact that $\left\|\boldsymbol{\Sigma}_i^{-1}\right\|_{\text{op},2} \leq \sigma^{-2}$. $\square$

Proof of Theorem 2. Recall that $\hat{\boldsymbol{\mu}}^* = \hat{\mu}(\mathbf{T}^*)$, $\mathbf{T}^* = (T_1^*, \ldots, T_{m^*}^*)^T$, the estimated FPCs $\hat{\xi}_k^* = \hat{\lambda}_k \hat{\phi}_k(\mathbf{T}^*)^T \hat{\boldsymbol{\Sigma}}^{*-1}(\mathbf{X}^* - \hat{\boldsymbol{\mu}}^*)$, $\hat{\boldsymbol{\Phi}}_K^*$ is analogous to $\hat{\boldsymbol{\Phi}}_{iK}$ while replacing the T_{ij} with T_j^*, and similarly for quantities such as $\boldsymbol{\Phi}_K^*$, $\hat{\boldsymbol{\Sigma}}^{*-1}$, and $\boldsymbol{\Sigma}^{*-1}$. Note that

$$\boldsymbol{\Sigma}_K^* - \hat{\boldsymbol{\Sigma}}_K^* = \boldsymbol{\Lambda}_K - \hat{\boldsymbol{\Lambda}}_K + \hat{\boldsymbol{\Lambda}}_K \hat{\boldsymbol{\Phi}}_K^{*T} \hat{\boldsymbol{\Sigma}}^{*-1} \hat{\boldsymbol{\Phi}}_K^* \hat{\boldsymbol{\Lambda}}_K - \boldsymbol{\Lambda}_K \boldsymbol{\Phi}_K^{*T} \boldsymbol{\Sigma}^{*-1} \boldsymbol{\Phi}_K^* \boldsymbol{\Lambda}_K, \tag{S.34}$$

where $\|\boldsymbol{\Lambda}_K - \hat{\boldsymbol{\Lambda}}_K\|_{\text{op},2} = O_p(a_n + b_n)$ follows from Theorem 5.2 in Zhang and Wang (2016) along with perturbation results (Bosq, 2000) and the fact that $\|\boldsymbol{\Lambda}_K - \hat{\boldsymbol{\Lambda}}_K\|_{\text{op},2} \leq \sqrt{K} \max_{1\leq k \leq K} |\lambda_k - \hat{\lambda}_k|$. Since $\hat{\lambda}_k \hat{\boldsymbol{\phi}}_k^* = \int_{\mathcal{T}} \hat{\Gamma}(\mathbf{T}^*, t) \hat{\phi}_k(t) dt$ and writing $\hat{\mathbf{e}}_k^* = \int_{\mathcal{T}} \hat{\Gamma}(\mathbf{T}^*, t) \hat{\phi}_k(t) dt - \hat{\boldsymbol{\Sigma}}^* \mathbf{W}^* \hat{\boldsymbol{\phi}}_k^*$, we have that the (j, l) entry of $\hat{\boldsymbol{\Lambda}}_K \hat{\boldsymbol{\Phi}}_K^{*T} \hat{\boldsymbol{\Sigma}}^{*-1} \hat{\boldsymbol{\Phi}}_K^* \hat{\boldsymbol{\Lambda}}_K$ is given by

$$\begin{aligned}
[\hat{\boldsymbol{\Lambda}}_K \hat{\boldsymbol{\Phi}}_K^{*T} \hat{\boldsymbol{\Sigma}}^{*-1} \hat{\boldsymbol{\Phi}}_K^* \hat{\boldsymbol{\Lambda}}_K]_{j,l} &= (\hat{\mathbf{e}}_j^{*T} \hat{\boldsymbol{\Sigma}}^{*-1} + \hat{\boldsymbol{\phi}}_j^{*T} \mathbf{W}^*)(\hat{\mathbf{e}}_l^* + \hat{\boldsymbol{\Sigma}}^* \mathbf{W}^* \hat{\boldsymbol{\phi}}_l^*) \\
&= \hat{\mathbf{e}}_j^{*T} \hat{\boldsymbol{\Sigma}}^{*-1} \hat{\mathbf{e}}_l^* + \hat{\mathbf{e}}_j^{*T} \mathbf{W}^* \hat{\boldsymbol{\phi}}_l^* + \hat{\boldsymbol{\phi}}_j^{*T} \mathbf{W}^* \hat{\mathbf{e}}_l^* + \hat{\boldsymbol{\phi}}_j^{*T} \mathbf{W}^* \hat{\boldsymbol{\Sigma}}^* \mathbf{W}^* \hat{\boldsymbol{\phi}}_l^*,
\end{aligned} \tag{S.35}$$

where $1 \leq j, l \leq K$. Denote by $\hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T})$ the matrix whose (i, j) element is $\hat{\Gamma}(T_i^*, T_j^*)$, $1 \leq i, j \leq m^*$, and similarly define $\Gamma(\mathbf{T}^*, \mathbf{T}^{*T})$. Also note that $\hat{\boldsymbol{\Sigma}}^* = \hat{\sigma}^2 I_{m^*} + \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T})$, where $I_{m^*} \in \mathbb{R}^{m^* \times m^*}$ is the identity matrix. From (S.21), (S.76), (S.87), (S.112), Lemma S2, and using that $\|\hat{\boldsymbol{\Sigma}}^{*-1} -$

$\boldsymbol{\Sigma}^{*-1}\|_{\mathrm{op},2} = O_p(m^*(a_n + b_n))$ along with the condition $m^*(a_n + b_n) = o(1)$ as $n \to \infty$, it follows that $\|\hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T}) - \Gamma(\mathbf{T}^*, \mathbf{T}^{*T})\|_2 = O_p(m^*(a_n + b_n))$, $\|\mathbf{W}^*(\hat{\boldsymbol{\phi}}_p^* - \boldsymbol{\phi}_p^*)\|_2 = O_p(m^{*-1/2}(a_n + b_n))$, $p = j, l$, $\|\Gamma(\mathbf{T}^*, \mathbf{T}^{*T})\|_{\mathrm{op},2} = O(m^*)$, $\|\boldsymbol{\Sigma}^*\|_{\mathrm{op},2} = O(m^*)$, $\|\hat{\boldsymbol{\Sigma}}^*\|_{\mathrm{op},2} = O_p(m^*)$, $\|\mathbf{W}^*\boldsymbol{\phi}_p^*\|_2 = O_p(m^{*-1/2})$, $p = j, l$, $\|\hat{\boldsymbol{\Sigma}}^* - \boldsymbol{\Sigma}^*\|_{\mathrm{op},2} = O_p(m^*(a_n + b_n))$, $\|\mathbf{W}^*\|_2 = O_p(m^{*-1/2})$, $\|\mathbf{e}_p^*\|_2 = O_p(m^{*-1/2})$ and $\|\hat{\mathbf{e}}_p^* - \mathbf{e}_p^*\|_2 = O_p(m^{*1/2}(a_n + b_n))$, $p = j, l$. These bounds imply

$$\begin{aligned}
\hat{\boldsymbol{\phi}}_j^{*T}\mathbf{W}^*\hat{\boldsymbol{\Sigma}}^*\mathbf{W}^*\hat{\boldsymbol{\phi}}_l^* - \boldsymbol{\phi}_j^{*T}\mathbf{W}^*\boldsymbol{\Sigma}^*\mathbf{W}^*\boldsymbol{\phi}_l^* &= O_p(a_n + b_n),\\
\hat{\mathbf{e}}_j^{*T}\hat{\boldsymbol{\Sigma}}^{*-1}\hat{\mathbf{e}}_l^* - \mathbf{e}_j^{*T}\boldsymbol{\Sigma}^{*-1}\mathbf{e}_l^* &= O_p(a_n + b_n),\\
\hat{\mathbf{e}}_j^{*T}\mathbf{W}^*\hat{\boldsymbol{\phi}}_l^* - \mathbf{e}_j^{*T}\mathbf{W}^*\boldsymbol{\phi}_l^* &= O_p(a_n + b_n),\\
\hat{\boldsymbol{\phi}}_j^{*T}\mathbf{W}^*\hat{\mathbf{e}}_l^* - \boldsymbol{\phi}_j^{*T}\mathbf{W}^*\mathbf{e}_l^* &= O_p(a_n + b_n),
\end{aligned}$$

which combined with (S.35) leads to

$$[\hat{\boldsymbol{\Lambda}}_K\hat{\boldsymbol{\Phi}}_K^{*T}\hat{\boldsymbol{\Sigma}}^{*-1}\hat{\boldsymbol{\Phi}}_K^*\hat{\boldsymbol{\Lambda}}_K]_{j,l} - [\boldsymbol{\Lambda}_K\boldsymbol{\Phi}_K^{*T}\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Phi}_K^*\boldsymbol{\Lambda}_K]_{j,l} = O_p(a_n + b_n).$$

Hence $\|\hat{\boldsymbol{\Lambda}}_K\hat{\boldsymbol{\Phi}}_K^{*T}\hat{\boldsymbol{\Sigma}}^{*-1}\hat{\boldsymbol{\Phi}}_K^*\hat{\boldsymbol{\Lambda}}_K - \boldsymbol{\Lambda}_K\boldsymbol{\Phi}_K^{*T}\boldsymbol{\Sigma}^{*-1}\boldsymbol{\Phi}_K^*\boldsymbol{\Lambda}_K\|_F = O_p(a_n + b_n)$ and the result follows from (S.34). □

Proof of Theorem 4. Let $\upsilon_K = \sum_{k=1}^K \lambda_k^{-1/2}\delta_k^{-1}$ and $\nu_K = \sum_{k=1}^K \lambda_k^{-1}$. Note that

$$\begin{aligned}
\|\hat{\mu}_K^* - \tilde{\mu}_K^*\|_{L^2} &= \|\hat{\boldsymbol{\xi}}_K^{*T}\hat{\boldsymbol{\Phi}}_K - \tilde{\boldsymbol{\xi}}_K^{*T}\boldsymbol{\Phi}_K\|_{L^2}\\
&\le \|(\hat{\boldsymbol{\xi}}_K^* - \tilde{\boldsymbol{\xi}}_K^*)^T(\hat{\boldsymbol{\Phi}}_K - \boldsymbol{\Phi}_K)\|_{L^2} + \|(\hat{\boldsymbol{\xi}}_K^* - \tilde{\boldsymbol{\xi}}_K^*)^T\boldsymbol{\Phi}_K\|_{L^2}\\
&\quad + \|\tilde{\boldsymbol{\xi}}_K^{*T}(\hat{\boldsymbol{\Phi}}_K - \boldsymbol{\Phi}_K)\|_{L^2}. \qquad \text{(S.36)}
\end{aligned}$$

Now, by the Cauchy–Schwarz inequality,

$$\|(\hat{\boldsymbol{\xi}}_K^* - \tilde{\boldsymbol{\xi}}_K^*)^T(\hat{\boldsymbol{\Phi}}_K - \boldsymbol{\Phi}_K)\|_{L^2} \le \|\hat{\boldsymbol{\xi}}_K^* - \tilde{\boldsymbol{\xi}}_K^*\|_2 \sum_{k=1}^{K} \|\hat{\phi}_k - \phi_k\|_{L^2} \lesssim \left(\sum_{k=1}^{K} \delta_k^{-1}\right) \|\hat{\boldsymbol{\xi}}_K^* - \tilde{\boldsymbol{\xi}}_K^*\|_2 \|\hat{\Xi} - \Xi\|_{\mathrm{op}}, \quad \text{(S.37)}$$

and by orthonormality of the ϕ_k,

$$\|(\hat{\boldsymbol{\xi}}_K^* - \tilde{\boldsymbol{\xi}}_K^*)^T \boldsymbol{\Phi}_K\|_{L^2} \le \|\hat{\boldsymbol{\xi}}_K^* - \tilde{\boldsymbol{\xi}}_K^*\|_2. \quad \text{(S.38)}$$

Also note that

$$E(\|\tilde{\boldsymbol{\xi}}_K^*\|_2^2) = \mathrm{trace}(E[E(\tilde{\boldsymbol{\xi}}_K^* \tilde{\boldsymbol{\xi}}_K^{*T} | \mathbf{T}^*)]) = E(\mathrm{trace}(\boldsymbol{\Lambda}_K \boldsymbol{\Phi}_K^{*T} \boldsymbol{\Sigma}^{*-1} \boldsymbol{\Phi}_K^* \boldsymbol{\Lambda}_K)) = E\left(\sum_{k=1}^{K} \lambda_k^2 \boldsymbol{\phi}_k^{*T} \boldsymbol{\Sigma}^{*-1} \boldsymbol{\phi}_k^*\right),$$

and

$$\lambda_j^2 \boldsymbol{\phi}_j^{*T} \boldsymbol{\Sigma}^{*-1} \boldsymbol{\phi}_j^* = \mathbf{e}_j^{*T} \boldsymbol{\Sigma}^{*-1} \mathbf{e}_j^* + 2\mathbf{e}_j^{*T} \mathbf{W}^* \boldsymbol{\phi}_j^* + \boldsymbol{\phi}_j^{*T} \mathbf{W}^* \boldsymbol{\Sigma}^* \mathbf{W}^* \boldsymbol{\phi}_j^*,$$

where $j = 1, \dots, K$. Similar arguments as the ones outlined in the proof of Theorem 3 then show that for large enough n

$$E(\|\tilde{\boldsymbol{\xi}}_K^*\|_2^2) = E\left(\sum_{k=1}^{K} \lambda_k^2 \boldsymbol{\phi}_k^{*T} \boldsymbol{\Sigma}^{*-1} \boldsymbol{\phi}_k^*\right) \lesssim m^{*(1-2\delta)} + m^{*-\delta} + \sum_{k=1}^{K} \lambda_k \lesssim m^{*(1-2\delta)} + \sum_{k=1}^{K} \lambda_k.$$

Since $\delta \in (1/2, 1)$ and $\sum_{k=1}^{\infty} \lambda_k < \infty$, this implies

$$\|\tilde{\boldsymbol{\xi}}_K^*\|_2 = O_p(1). \quad \text{(S.39)}$$

Observing

$$\|\tilde{\boldsymbol{\xi}}_K^{*T}(\hat{\boldsymbol{\Phi}}_K - \boldsymbol{\Phi}_K)\|_{L^2} \le \|\tilde{\boldsymbol{\xi}}_K^*\|_2 \sum_{k=1}^{K} \|\hat{\phi}_k - \phi_k\|_{L^2} \lesssim \left(\sum_{k=1}^{K} \delta_k^{-1}\right) \|\hat{\Xi} - \Xi\|_{\mathrm{op}} \|\tilde{\boldsymbol{\xi}}_K^*\|_2,$$

and using (S.77) along with (S.39) leads to

$$\|\tilde{\boldsymbol{\xi}}_K^{*T}(\hat{\boldsymbol{\Phi}}_K - \boldsymbol{\Phi}_K)\|_{L^2} = O_p\left((a_n + b_n)\sum_{k=1}^K \delta_k^{-1}\right). \tag{S.40}$$

In view of (S.36), (S.37), (S.38), (S.40), the condition $\upsilon_K(a_n+b_n) = o(1)$ which implies $(a_n + b_n)\sum_{k=1}^K \delta_k^{-1} = o(1)$ as $n \to \infty$, and employing Lemma S3 leads to

$$\begin{aligned}
&\|\hat{\mu}_K^* - \tilde{\mu}_K^*\|_{L^2} \\
&= O_p\Bigg((a_n + b_n)\left(\sum_{k=1}^K \delta_k^{-1}\right) + m^{*1/2}(a_n + b_n)\left(\sum_{k=1}^K \delta_k^{-2}\lambda_k^{-2}\right)^{1/2} + m^{*-1/2}\left(\sum_{k=1}^K \lambda_k^{-2}\right)^{1/2} \\
&\quad + m^*(a_n + b_n)\left(\sum_{k=1}^K \lambda_k^{-2}\right)^{1/2} + m^{*2}(a_n + b_n)^2\left(\sum_{k=1}^K \delta_k^{-2}\lambda_k^{-2}\right)^{1/2}\Bigg).
\end{aligned} \tag{S.41}$$

Observe

$$\mathcal{W}_2^2(\hat{\mathcal{G}}_K^*, \mathcal{A}_{X^{*c}}) \le E(\|g_1 - g_2\|_{L^2}^2 \mid (\mathbf{X}_j, \mathbf{T}_j)_{j=0}^n),$$

where $\mathbf{X}_0 := \mathbf{X}^*$ and $\mathbf{T}_0 := \mathbf{T}^*$, the random element $g_1 \in L^2$ has conditional distribution $g_1 \sim \hat{\mathcal{G}}_K^*$ given $(\mathbf{X}_j, \mathbf{T}_j)_{j=0}^n$, and $g_2(\cdot) = X^{*c}(\cdot)$ almost surely. Since $E(g_1 \mid (\mathbf{X}_j, \mathbf{T}_j)_{j=0}^n) = \hat{\mu}_K^*$ and $\mathrm{Var}(g_1(t) \mid (\mathbf{X}_j, \mathbf{T}_j)_{j=0}^n) = \hat{\Gamma}_K^*(t,t)$, $t \in \mathcal{T}$, we obtain

$$\begin{aligned}
\mathcal{W}_2^2(\hat{\mathcal{G}}_K^*, \mathcal{A}_{X^{*c}}) &\le E(\|g_1 - \hat{\mu}_K^*\|_{L^2}^2 \mid (\mathbf{X}_j, \mathbf{T}_j)_{j=0}^n) + \|\hat{\mu}_K^* - X^{*c}\|_{L^2}^2 \\
&= \int_{\mathcal{T}} \hat{\Gamma}_K^*(t,t)dt + \|\hat{\mu}_K^* - X^{*c}\|_{L^2}^2 \\
&\le \int_{\mathcal{T}} (\hat{\Gamma}_K^*(t,t) - \Gamma_K^*(t,t))dt + \|\hat{\mu}_K^* - X^{*c}\|_{L^2}^2 + O_p(m^{*(1-2\delta)}),
\end{aligned}$$

where the equality follows from Fubini's Theorem and the last inequality is due to $\int_{\mathcal{T}} \Gamma_K^*(t,t)dt = O_p(m^{1-2\delta})$, which follows analogously as in (S.30). Combining (S.41) and arguments analogous to those in the proof of Theorem 3 lead to

$$
\begin{aligned}
&\|\hat{\mu}_K^* - X^{*c}\|_{L^2} \\
&\le \|\tilde{\mu}_K^* - X^{*c}\|_{L^2} + \|\hat{\mu}_K^* - \tilde{\mu}_K^*\|_{L^2} \\
&= O_p\Big[m^{*(1/2-\delta)} + \left(\sum_{k=K+1}^{\infty} \lambda_k\right)^{1/2} + (a_n + b_n)\left(\sum_{k=1}^{K} \delta_k^{-1}\right) \\
&\qquad + m^{*1/2}(a_n + b_n)\left(\sum_{k=1}^{K} \delta_k^{-2}\lambda_k^{-2}\right)^{1/2} + m^{*-1/2}\left(\sum_{k=1}^{K} \lambda_k^{-2}\right)^{1/2} \\
&\qquad + m^{*}(a_n + b_n)\left(\sum_{k=1}^{K} \lambda_k^{-2}\right)^{1/2} + m^{*2}(a_n + b_n)^2\left(\sum_{k=1}^{K} \delta_k^{-2}\lambda_k^{-2}\right)^{1/2}\Big].
\end{aligned}
$$

From Lemma S4 we have

$$
\begin{aligned}
\int_{\mathcal{T}} (\hat{\Gamma}_K^*(t,t) - \Gamma_K^*(t,t))dt &= \text{trace}(\hat{\mathbf{\Sigma}}_K^* - \mathbf{\Sigma}_K^*) \\
&= O_p\left(m^*(a_n + b_n)^2 \sum_{k=1}^{K} \lambda_k^{-2}\delta_k^{-2} + (a_n + b_n)\sum_{k=1}^{K} \lambda_k^{-2}\delta_k^{-1}\right).
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&\mathcal{W}_2^2(\hat{\mathcal{G}}_K^*, \mathcal{A}_{X^{*c}}) \\
&= O_p\Big[m^{*(1-2\delta)} + \sum_{k=K+1}^{\infty} \lambda_k + (a_n + b_n)^2\left(\sum_{k=1}^{K} \delta_k^{-1}\right)^2 + m^*(a_n + b_n)^2 \sum_{k=1}^{K} \delta_k^{-2}\lambda_k^{-2} \\
&\qquad + m^{*-1}\sum_{k=1}^{K} \lambda_k^{-2} + m^{*2}(a_n + b_n)^2 \sum_{k=1}^{K} \lambda_k^{-2} + m^{*4}(a_n + b_n)^4 \sum_{k=1}^{K} \delta_k^{-2}\lambda_k^{-2} \\
&\qquad + (a_n + b_n)\sum_{k=1}^{K} \lambda_k^{-2}\delta_k^{-1}\Big],
\end{aligned}
$$

---

and the result follows. □

Proof of Theorem 5. We use the fact that for a normal random variable $Z_1 \sim N(\kappa_1, \kappa_2^2)$ and $t \in (0,1)$ it holds that $Q_1(t) = \kappa_2 q(t) + \kappa_1$, where $Q_1(\cdot)$ and $q(\cdot)$ are the quantile functions corresponding to Z_1 and a standard normal random variate, respectively. Note that since $|\lambda_{\min}(\hat{\mathbf{\Sigma}}_{iK}) - \lambda_{\min}(\mathbf{\Sigma}_{iK})| \leq \left\| \hat{\mathbf{\Sigma}}_{iK} - \mathbf{\Sigma}_{iK} \right\|_{\text{op},2} = o_p(1)$, where the $o_p(1)$ term is uniform in i (see the proof of Lemma S12), and $\lambda_{\min}(\mathbf{\Sigma}_{iK}) \geq \kappa_0$ a.s., we have

$$\begin{aligned}
P\left(\left\| \hat{\mathbf{\Sigma}}_{iK} - \mathbf{\Sigma}_{iK} \right\|_{\text{op},2} \leq \kappa_0/2 \right) &= P\left(\kappa_0 - \left\| \hat{\mathbf{\Sigma}}_{iK} - \mathbf{\Sigma}_{iK} \right\|_{\text{op},2} \geq \kappa_0/2 \right) \\
&\leq P\left(\lambda_{\min}(\mathbf{\Sigma}_{iK}) - \left\| \hat{\mathbf{\Sigma}}_{iK} - \mathbf{\Sigma}_{iK} \right\|_{\text{op},2} \geq \kappa_0/2 \right) \\
&\leq P\left(\lambda_{\min}(\hat{\mathbf{\Sigma}}_{iK}) \geq \kappa_0/2 \right),
\end{aligned}$$

which implies $\lambda_{\min}(\hat{\mathbf{\Sigma}}_{iK}) \geq \kappa_0/2$ with probability tending to 1. For the remainder of the proof we work on this event. From the closed form expression for the 2-Wasserstein distance between one-dimensional distributions with finite second moments,

$$\begin{aligned}
\mathcal{W}_2^2(\tilde{\mathcal{P}}_{iK}, \mathcal{P}_{iK}) &= \int_0^1 \left([(\boldsymbol{\beta}_K^T \hat{\mathbf{\Sigma}}_{iK} \boldsymbol{\beta}_K)^{1/2} - (\boldsymbol{\beta}_K^T \mathbf{\Sigma}_{iK} \boldsymbol{\beta}_K)^{1/2}] q(t) + \boldsymbol{\beta}_K^T (\hat{\boldsymbol{\xi}}_{iK} - \tilde{\boldsymbol{\xi}}_{iK}) \right)^2 dt \\
&= [(\boldsymbol{\beta}_K^T \hat{\mathbf{\Sigma}}_{iK} \boldsymbol{\beta}_K)^{1/2} - (\boldsymbol{\beta}_K^T \mathbf{\Sigma}_{iK} \boldsymbol{\beta}_K)^{1/2}]^2 \int_0^1 q^2(t) dt + (\boldsymbol{\beta}_K^T (\hat{\boldsymbol{\xi}}_{iK} - \tilde{\boldsymbol{\xi}}_{iK}))^2 \\
&\quad + 2[(\boldsymbol{\beta}_K^T \hat{\mathbf{\Sigma}}_{iK} \boldsymbol{\beta}_K)^{1/2} - (\boldsymbol{\beta}_K^T \mathbf{\Sigma}_{iK} \boldsymbol{\beta}_K)^{1/2}]\, \boldsymbol{\beta}_K^T (\hat{\boldsymbol{\xi}}_{iK} - \tilde{\boldsymbol{\xi}}_{iK}) \int_0^1 q(t) dt \\
&\leq \frac{(\boldsymbol{\beta}_K^T (\hat{\mathbf{\Sigma}}_{iK} - \mathbf{\Sigma}_{iK}) \boldsymbol{\beta}_K)^2}{\boldsymbol{\beta}_K^T \mathbf{\Sigma}_{iK} \boldsymbol{\beta}_K} \int_0^1 q^2(t) dt + (\boldsymbol{\beta}_K^T (\hat{\boldsymbol{\xi}}_{iK} - \tilde{\boldsymbol{\xi}}_{iK}))^2,
\end{aligned} \tag{S.42}$$

where the last inequality follows from the fact that $\int_0^1 q(t)dt = E(Z) = 0$, where $Z \sim N(0,1)$, and using the inequality $(\sqrt{x} - \sqrt{y})^2 \leq (x-y)^2/y$ which is valid for any scalars $x \geq 0$ and $y > 0$. Since $\int_0^1 q^2(t)dt = E(Z^2) < \infty$, it then suffices to control the terms $\boldsymbol{\beta}_K^T(\hat{\boldsymbol{\Sigma}}_{iK} - \boldsymbol{\Sigma}_{iK})\boldsymbol{\beta}_K$ and $(\boldsymbol{\beta}_K^T(\hat{\boldsymbol{\xi}}_{iK} - \tilde{\boldsymbol{\xi}}_{iK}))^2$. From the proof of Lemma S12, we have $\left\|\boldsymbol{\Sigma}_{iK} - \hat{\boldsymbol{\Sigma}}_{iK}\right\|_F = O(a_n + b_n)$ a.s. as $n \to \infty$, where the $O(a_n + b_n)$ term is uniform over i, and similar arguments as in the proof of Theorem 2 in Dai *et al.* (2018) show that $|\hat{\xi}_{ik} - \tilde{\xi}_{ik}| = O(a_n + b_n)\|\mathbf{X}_i - \boldsymbol{\mu}_i\|_2 = O(a_n + b_n)O_p(1) = O_p(a_n + b_n)$, $k = 1, \ldots, K$. Thus, $(\boldsymbol{\beta}_K^T(\hat{\boldsymbol{\xi}}_{iK} - \tilde{\boldsymbol{\xi}}_{iK}))^2 \leq \|\boldsymbol{\beta}_K\|_2^2 \|\hat{\boldsymbol{\xi}}_{iK} - \tilde{\boldsymbol{\xi}}_{iK}\|_2^2 = O_p((a_n + b_n)^2)$ and properties of the operator norm show that $|\boldsymbol{\beta}_K^T(\hat{\boldsymbol{\Sigma}}_{iK} - \boldsymbol{\Sigma}_{iK})\boldsymbol{\beta}_K| \leq \|\boldsymbol{\beta}_K\|_2^2 \left\|\hat{\boldsymbol{\Sigma}}_{iK} - \boldsymbol{\Sigma}_{iK}\right\|_F = O(a_n + b_n)$ a.s. as $n \to \infty$. This along with (S.42) leads to

$$\mathcal{W}_2(\tilde{\mathcal{P}}_{iK}, \mathcal{P}_{iK}) = O_p(a_n + b_n). \tag{S.43}$$

Similar arguments show that

$$\mathcal{W}_2^2(\hat{\mathcal{P}}_{iK}, \tilde{\mathcal{P}}_{iK}) \leq \frac{(\hat{\boldsymbol{\beta}}_K^T \hat{\boldsymbol{\Sigma}}_{iK} \hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K^T \hat{\boldsymbol{\Sigma}}_{iK} \boldsymbol{\beta}_K)^2}{\boldsymbol{\beta}_K^T \hat{\boldsymbol{\Sigma}}_{iK} \boldsymbol{\beta}_K} \int_0^1 q^2(t)dt + ((\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K)^T \hat{\boldsymbol{\xi}}_{iK} + \hat{\beta}_0 - \beta_0)^2, \tag{S.44}$$

and

$$\begin{aligned}
&|\hat{\boldsymbol{\beta}}_K^T \hat{\boldsymbol{\Sigma}}_{iK} \hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K^T \hat{\boldsymbol{\Sigma}}_{iK} \boldsymbol{\beta}_K| \\
&= |(\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K)^T \hat{\boldsymbol{\Sigma}}_{iK} \hat{\boldsymbol{\beta}}_K + \boldsymbol{\beta}_K^T \hat{\boldsymbol{\Sigma}}_{iK} (\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K)| \\
&\leq \left\|\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K\right\|_2^2 \left\|\hat{\boldsymbol{\Sigma}}_{iK} - \boldsymbol{\Sigma}_{iK}\right\|_{\text{op},2} + \left\|\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K\right\|_2 \left\|\hat{\boldsymbol{\Sigma}}_{iK} - \boldsymbol{\Sigma}_{iK}\right\|_{\text{op},2} \|\boldsymbol{\beta}_K\|_2 \\
&\quad + \left\|\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K\right\|_2^2 \|\boldsymbol{\Sigma}_{iK}\|_{\text{op},2} + \left\|\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K\right\|_2 \|\boldsymbol{\Sigma}_{iK}\|_{\text{op},2} \|\boldsymbol{\beta}_K\|_2 \\
&= O_p(\alpha_n),
\end{aligned} \tag{S.45}$$

where the first inequality follows from properties of the operator norm; the last equality is due to Lemma S11 and the facts that $h \asymp n^{-1/3}$ implies that the rate $\tau_M\left[\left(\frac{1}{nh}+h^2\right)^{1/2}+a_n\right]$ is faster than $c_n\upsilon_M$, $\left\|\hat{\boldsymbol{\Sigma}}_{iK}-\boldsymbol{\Sigma}_{iK}\right\|_F = O(a_n+b_n)$ a.s. as $n\to\infty$ and that $\|\boldsymbol{\Sigma}_{iK}\|_{\text{op},2}$ is uniformly bounded in i in the sparse case. Since $|\lambda_{\min}(\hat{\boldsymbol{\Sigma}}_{iK})-\lambda_{\min}(\boldsymbol{\Sigma}_{iK})| \le \left\|\hat{\boldsymbol{\Sigma}}_{iK}-\boldsymbol{\Sigma}_{iK}\right\|_{\text{op},2}$, we have

$$\begin{aligned}
\boldsymbol{\beta}_K^T\hat{\boldsymbol{\Sigma}}_{iK}\boldsymbol{\beta}_K &\ge \boldsymbol{\beta}_K^T\boldsymbol{\beta}_K\lambda_{\min}(\hat{\boldsymbol{\Sigma}}_{iK})\\
&\ge \boldsymbol{\beta}_K^T\boldsymbol{\beta}_K(\lambda_{\min}(\boldsymbol{\Sigma}_{iK})-\left\|\hat{\boldsymbol{\Sigma}}_{iK}-\boldsymbol{\Sigma}_{iK}\right\|_{\text{op},2})1_{\{\lambda_{\min}(\boldsymbol{\Sigma}_{iK})\ge\left\|\hat{\boldsymbol{\Sigma}}_{iK}-\boldsymbol{\Sigma}_{iK}\right\|_{\text{op},2}\}}.
\end{aligned}$$

Thus, using that $\left\|\hat{\boldsymbol{\Sigma}}_{iK}-\boldsymbol{\Sigma}_{iK}\right\|_{\text{op},2} = o_p(1)$, where the $o_p(1)$ term is uniform in i, $\lambda_{\min}(\boldsymbol{\Sigma}_{iK}) \ge \kappa_0$ a.s., and writing

$$p_0 = P\left[\frac{1}{\boldsymbol{\beta}_K^T\hat{\boldsymbol{\Sigma}}_{iK}\boldsymbol{\beta}_K} \le \frac{2}{\boldsymbol{\beta}_K^T\boldsymbol{\beta}_K\lambda_{\min}(\boldsymbol{\Sigma}_{iK})} \text{ and } \lambda_{\min}(\hat{\boldsymbol{\Sigma}}_{iK}) \ge \kappa_0/2\right],$$

it follows that

$$\begin{aligned}
p_0 &\ge P[\boldsymbol{\beta}_K^T\boldsymbol{\beta}_K\lambda_{\min}(\boldsymbol{\Sigma}_{iK}) \le 2\boldsymbol{\beta}_K^T\boldsymbol{\beta}_K\lambda_{\min}(\hat{\boldsymbol{\Sigma}}_{iK}) \text{ and } \lambda_{\min}(\hat{\boldsymbol{\Sigma}}_{iK}) \ge \kappa_0/2]\\
&\ge P[\boldsymbol{\beta}_K^T\boldsymbol{\beta}_K\lambda_{\min}(\boldsymbol{\Sigma}_{iK}) \le 2\boldsymbol{\beta}_K^T\boldsymbol{\beta}_K(\lambda_{\min}(\boldsymbol{\Sigma}_{iK})-\left\|\hat{\boldsymbol{\Sigma}}_{iK}-\boldsymbol{\Sigma}_{iK}\right\|_{\text{op},2}) \text{ and } \lambda_{\min}(\hat{\boldsymbol{\Sigma}}_{iK}) \ge \kappa_0/2]\\
&\ge P[\kappa_0/2 \ge \left\|\hat{\boldsymbol{\Sigma}}_{iK}-\boldsymbol{\Sigma}_{iK}\right\|_{\text{op},2} \text{ and } \lambda_{\min}(\hat{\boldsymbol{\Sigma}}_{iK}) \ge \kappa_0/2]\\
&\ge 1-P\left[\left\|\hat{\boldsymbol{\Sigma}}_{iK}-\boldsymbol{\Sigma}_{iK}\right\|_{\text{op},2} > \kappa_0/2\right]-P[\lambda_{\min}(\hat{\boldsymbol{\Sigma}}_{iK}) < \kappa_0/2].
\end{aligned}$$

This implies $p_0\to 1$ as $n\to\infty$ and hence the event $(\boldsymbol{\beta}_K^T\hat{\boldsymbol{\Sigma}}_{iK}\boldsymbol{\beta}_K)^{-1} \le 2(\boldsymbol{\beta}_K^T\boldsymbol{\beta}_K\lambda_{\min}(\boldsymbol{\Sigma}_{iK}))^{-1}$ with $\lambda_{\min}(\hat{\boldsymbol{\Sigma}}_{iK}) \ge \kappa_0/2$ occurs with probability tending to 1. It then suffices to work on this event in what follows. Combining with (S.44), (S.45), and

$$\begin{aligned}
|(\hat{\boldsymbol{\beta}}_K-\boldsymbol{\beta}_K)^T\hat{\boldsymbol{\xi}}_{iK}+\hat{\beta}_0-\beta_0| &\le \left\|\hat{\boldsymbol{\beta}}_K-\boldsymbol{\beta}_K\right\|_2\left(\left\|\hat{\boldsymbol{\xi}}_{iK}-\tilde{\boldsymbol{\xi}}_{iK}\right\|_2+\left\|\tilde{\boldsymbol{\xi}}_{iK}\right\|_2\right)+|\hat{\beta}_0-\beta_0|\\
&= O_p(\alpha_n),
\end{aligned}$$

which follows from Lemma S11 and the facts that $\hat{\beta}_0 - \beta_0 = \bar{Y}_n - E(Y) = O_p(n^{-1/2})$, $\left\|\hat{\boldsymbol{\xi}}_{iK} - \tilde{\boldsymbol{\xi}}_{iK}\right\|_2 = O_p(a_n + b_n)$ and $\left\|\tilde{\boldsymbol{\xi}}_{iK}\right\|_2 = O_p(1)$ hold uniformly in i, then leads to

$$\mathcal{W}_2(\hat{\mathcal{P}}_{iK}, \tilde{\mathcal{P}}_{iK}) = O_p(\alpha_n). \tag{S.46}$$

The result in (4.17) then follows from (S.43) and (S.46).

Denote by φ and Φ the density and cdf of a standard normal random variable, and define the quantities $\tilde{u}_{in} = \beta_0 + \boldsymbol{\beta}_K^T \hat{\boldsymbol{\xi}}_{iK}$, $\tilde{\sigma}_{in} = (\boldsymbol{\beta}_K^T \hat{\boldsymbol{\Sigma}}_{iK} \boldsymbol{\beta}_K)^{1/2}$, $u_i = \beta_0 + \boldsymbol{\beta}_K^T \tilde{\boldsymbol{\xi}}_{iK}$, $\sigma_i = (\boldsymbol{\beta}_K^T \boldsymbol{\Sigma}_{iK} \boldsymbol{\beta}_K)^{1/2}$ and $\Delta_{in}(t) = (t - u_i)/\sigma_i - (t - \tilde{u}_{in})/\tilde{\sigma}_{in}$, $t \in \mathbb{R}$. Then

$$\sup_{t\in\mathbb{R}} |\tilde{F}_{iK}(t) - F_{iK}(t)| = \sup_{t\in\mathbb{R}} \left|\Phi\left(\frac{t - \tilde{u}_{in}}{\tilde{\sigma}_{in}}\right) - \Phi\left(\frac{t - u_i}{\sigma_i}\right)\right| = \sup_{t\in\mathbb{R}} \left|\varphi(\varepsilon_s)\Delta_{in}(t)\right|, \tag{S.47}$$

where the second equality follows by a Taylor expansion and ε_s is between $(t - \tilde{u}_{in})/\tilde{\sigma}_{in}$ and $(t - \mu_i)/\sigma_i$. Defining $r_{in}(t) = (t - \tilde{u}_{in})/\tilde{\sigma}_{in}$, $r_i(t) = (t - u_i)/\sigma_i$ and setting $I_{in} = [\min\{u_i, \tilde{u}_{in}\}, \max\{u_i, \tilde{u}_{in}\}]$,

$$\begin{aligned} \left|\varphi(\varepsilon_s)\Delta_{in}(t)\right| &\le \varphi(0)\left|\Delta_{in}(t)\right| 1_{\{t\in I_{in}\}} + \varphi(\min\{|r_{in}(t)|, |r_i(t)|\})\left|\Delta_{in}(t)\right| 1_{\{t \in I_{in}^c\}} \\ &\le \varphi(0)\left|\Delta_{in}(t)\right| 1_{\{t\in I_{in}\}} + [\varphi(r_{in}(t)) + \varphi(r_i(t))]\left|\Delta_{in}(t)\right|. \end{aligned} \tag{S.48}$$

Since $\tilde{u}_{in} - u_i = O_p(a_n + b_n)$, $|\tilde{\sigma}_{in} - \sigma_i| \le |\tilde{\sigma}_{in}^2 - \sigma_i^2|/\sigma_i = O_p(a_n + b_n)$, $|\tilde{\sigma}_{in}^{-1} - \sigma_i^{-1}| \le |\tilde{\sigma}_{in} - \sigma_i|/(\tilde{\sigma}_{in}\sigma_i) \le |\tilde{\sigma}_{in} - \sigma_i|\sqrt{2}(\boldsymbol{\beta}_K^T \boldsymbol{\beta}_K \lambda_{\min}(\boldsymbol{\Sigma}_{iK}))^{-1/2}\sigma_i^{-1}$

and $\lambda_{\min}(\mathbf{\Sigma}_{iK}) \geq \kappa_0$ a.s., it follows that

$$\begin{aligned}|\Delta_{in}(t)| &= |(t-u_i)/\sigma_i - (t-\tilde{u}_{in})/\tilde{\sigma}_{in}| \\ &\leq \frac{1}{\sigma_i}|\tilde{u}_{in} - u_i| + |t-u_i|\left|\frac{1}{\tilde{\sigma}_{in}} - \frac{1}{\sigma_i}\right| + |\tilde{u}_{in} - u_i|\left|\frac{1}{\tilde{\sigma}_{in}} - \frac{1}{\sigma_i}\right| \\ &= O_p(a_n + b_n) + O_p(a_n + b_n)|t - u_i|, \end{aligned} \tag{S.49}$$

where both $O_p(a_n + b_n)$ terms are uniform in t. This implies

$$\sup_{t\in\mathbb{R}} \left|\Delta_{in}(t)\right| 1_{\{t\in I_{in}\}} \leq O_p(a_n + b_n) + O_p(a_n + b_n)|\tilde{u}_{in} - u_i| = O_p(a_n + b_n). \tag{S.50}$$

Since $\|\mathbf{\Sigma}_{iK}\|_{op}$ is uniformly bounded above in the sparse case, it is easy to show that $\varphi(r_i(t))|t - u_i| \leq O(1)$, where the $O(1)$ term is uniform in both t and i. This combined with (S.49) leads to

$$\sup_{t\in\mathbb{R}} \varphi(r_i(t))|\Delta_{in}(t)| = O_p(a_n + b_n). \tag{S.51}$$

From (S.49),

$$\sup_{t\in\mathbb{R}} \varphi(r_{in}(t))|\Delta_{in}(t)| \leq O_p(a_n + b_n) + O_p(a_n + b_n) \sup_{t\in\mathbb{R}} \varphi(r_{in}(t))|t - u_i|,$$

and the result then follows from (S.47), (S.48), (S.50) and (S.51) if we can show that $\varphi(r_{in}(t))|t - u_i| = O_p(1)$ uniformly in t. It is easy to see that

$$\begin{aligned}\varphi(r_{in}(t))|t - u_i| &\leq \varphi(r_{in}(t_1^*))(t_1^* - u_i)1_{\{t\geq u_i\}} + \varphi(r_{in}(t_2^*))(u_i - t_2^*)1_{\{t\leq u_i\}} \\ &\leq \varphi(r_{in}(t_1^*))(t_1^* - u_i) + \varphi(r_{in}(t_2^*))(u_i - t_2^*) \\ &\leq \varphi(0)(t_1^* - t_2^*),\end{aligned}$$

where $t_1^* = (u_i+\tilde{u}_{in}+\sqrt{(u_i-\tilde{u}_{in})^2+4\tilde{\sigma}_{in}^2})/2$ and $t_2^* = (u_i+\tilde{u}_{in}-\sqrt{(u_i-\tilde{u}_{in})^2+4\tilde{\sigma}_{in}^2})/2$. Since $\tilde{\sigma}_{in}$ is uniformly upper bounded in the sparse setting and $\tilde{u}_{in} - u_i = O_p(a_n + b_n)$, we obtain

$$\sup_{t\in\mathbb{R}} \varphi(r_{in}(t))|t-u_i| \le \varphi(0)\sqrt{(u_i-\tilde{u}_{in})^2+4\tilde{\sigma}_{in}^2} = O_p(1).$$

Therefore

$$\sup_{t\in\mathbb{R}}|\tilde{F}_{iK}(t) - F_{iK}(t)| = O_p(a_n+b_n), \tag{S.52}$$

so that it then remains to control the term $\sup_{t\in\mathbb{R}}|\hat{F}_{iK}(t) - \tilde{F}_{iK}(t)|$. For this purpose, define auxiliary quantities $\hat{u}_{in} = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_K^T\hat{\boldsymbol{\xi}}_{iK}$, $\hat{\sigma}_{in} = (\hat{\boldsymbol{\beta}}_K^T\hat{\boldsymbol{\Sigma}}_{iK}\hat{\boldsymbol{\beta}}_K)^{1/2}$ and $\hat{\Delta}_{in}(t) = (t-\hat{u}_{in})/\hat{\sigma}_{in} - (t-\tilde{u}_{in})/\tilde{\sigma}_{in}$, $t\in\mathbb{R}$. From Lemma S11 it follows that $\hat{u}_{in}-\tilde{u}_{in} = \hat{\beta}_0-\beta_0+(\hat{\boldsymbol{\beta}}_K-\boldsymbol{\beta}_K)^T(\hat{\boldsymbol{\xi}}_{iK}-\boldsymbol{\xi}_{iK})+(\hat{\boldsymbol{\beta}}_K-\boldsymbol{\beta}_K)^T\boldsymbol{\xi}_{iK} = O_p(\alpha_n)$, $|\hat{\sigma}_{in}-\tilde{\sigma}_{in}| \le |\hat{\sigma}_{in}^2-\tilde{\sigma}_{in}^2|/\tilde{\sigma}_{in} = O_p(\alpha_n)$, which is due to (S.45) and since $\tilde{\sigma}_{in}^{-1} \le \sqrt{2}(\boldsymbol{\beta}_K^T\boldsymbol{\beta}_K\lambda_{\min}(\boldsymbol{\Sigma}_{iK}))^{-1/2}$. Also, from (S.45) and using $\lambda_{\min}(\boldsymbol{\Sigma}_{iK}) \ge \kappa_0$ a.s. we have $|\hat{\sigma}_{in}-\tilde{\sigma}_{in}| \le |\hat{\sigma}_{in}^2-\tilde{\sigma}_{in}^2|/\tilde{\sigma}_{in} \le |\hat{\sigma}_{in}^2-\tilde{\sigma}_{in}^2|\sqrt{2}(\boldsymbol{\beta}_K^T\boldsymbol{\beta}_K\lambda_{\min}(\boldsymbol{\Sigma}_{iK}))^{-1/2} = o_p(1)$ and then $|\hat{\sigma}_{in}-\sigma_i| \le |\hat{\sigma}_{in}-\tilde{\sigma}_{in}| + |\tilde{\sigma}_{in}-\sigma_i| = o_p(1)$. This along with the fact that $\hat{\sigma}_{in} \ge \|\hat{\boldsymbol{\beta}}_K\|_2\kappa_0/2 \ge \|\boldsymbol{\beta}_K\|_2\kappa_0/4$ holds with probability tending to 1 implies $\hat{\sigma}_{in}^{-1} \le 2\sigma_i^{-1}$ with probability tending to 1 as $n\to\infty$. Combining this with $\lambda_{\min}(\boldsymbol{\Sigma}_{iK}) \ge \kappa_0$ a.s. then leads to

$$\Big|\frac{1}{\hat{\sigma}_{in}} - \frac{1}{\tilde{\sigma}_{in}}\Big| = O_p(\alpha_n),$$

where the bound is uniform in i, and similarly as in (S.49) we obtain

$$\begin{aligned}|\hat{\Delta}_{in}(t)| &\le |t-\tilde{u}_{in}|\,\Big|\frac{1}{\hat{\sigma}_{in}} - \frac{1}{\tilde{\sigma}_{in}}\Big| + |\hat{u}_{in}-\tilde{u}_{in}|\,\Big|\frac{1}{\hat{\sigma}_{in}} - \frac{1}{\tilde{\sigma}_{in}}\Big| + |\hat{u}_{in}-\tilde{u}_{in}|\frac{1}{\tilde{\sigma}_{in}} \\ &\le O_p(\alpha_n) + O_p(\alpha_n)|t-\tilde{u}_{in}|. \end{aligned} \tag{S.53}$$

Next

$$\varphi(r_{in}(t))|t-\tilde{u}_{in}|$$
$$\leq \varphi(1)\sqrt{\boldsymbol{\beta}_K^T\hat{\boldsymbol{\Sigma}}_{iK}\boldsymbol{\beta}_K} \leq \varphi(1)(\boldsymbol{\beta}_K^T\boldsymbol{\beta}_K)^{1/2}\left(\left\|\hat{\boldsymbol{\Sigma}}_{iK}-\boldsymbol{\Sigma}_{iK}\right\|_{\text{op},2}+\|\boldsymbol{\Sigma}_{iK}\|_{\text{op},2}\right)^{1/2}$$
$$=O_p(1),$$

where the $O_p(1)$ term is uniform in both t and i. This combined with (S.53) shows that

$$\sup_{t\in\mathbb{R}}\varphi(r_{in}(t))|\hat{\Delta}_{in}(t)| = O_p(\alpha_n). \tag{S.54}$$

Setting $\hat{r}_{in}(t)=(t-\hat{u}_{in})/\hat{\sigma}_{in}$, similar arguments as before lead to

$$\sup_{t\in\mathbb{R}}\varphi(\hat{r}_{in}(t))|t-\tilde{u}_{in}| \leq \varphi(0)\sqrt{(\hat{u}_{in}-\tilde{u}_{in})^2+4\hat{\sigma}_{in}^2}=O_p(1),$$

where the last equality is due to $|\hat{u}_{in}-\tilde{u}_{in}|=O_p(\alpha_n)$ and $\hat{\sigma}_{in}^2 \leq \boldsymbol{\beta}_K^T\boldsymbol{\beta}_K\left(\left\|\hat{\boldsymbol{\Sigma}}_{iK}-\boldsymbol{\Sigma}_{iK}\right\|_{\text{op},2}+\|\boldsymbol{\Sigma}_{iK}\|_{\text{op},2}\right)=O_p(1)$. With (S.53) this implies

$$\sup_{t\in\mathbb{R}}\varphi(\hat{r}_{in}(t))|\hat{\Delta}_{in}(t)| = O_p(\alpha_n). \tag{S.55}$$

Setting $\hat{I}_{in}=[\min\{\hat{u}_{in},\tilde{u}_{in}\},\max\{\hat{u}_{in},\tilde{u}_{in}\}]$, then similar arguments as the ones outlined in (S.47) and (S.48) shows that

$$\sup_{t\in\mathbb{R}}|\hat{F}_{iK}(t)-\tilde{F}_{iK}(t)| \leq \varphi(0)\left|\hat{\Delta}_{in}(t)\right|1_{\{t\in\hat{I}_{in}\}}+[\varphi(\hat{r}_{in}(t))+\varphi(r_{in}(t))]\left|\hat{\Delta}_{in}(t)\right|. \tag{S.56}$$

This together with $\sup_{t\in\mathbb{R}}\left|\hat{\Delta}_{in}(t)\right|1_{\{t\in\hat{I}_{in}\}} \leq O_p(\alpha_n)+O_p(\alpha_n)|\hat{u}_{in}-\tilde{u}_{in}| = O_p(\alpha_n)$, where the latter follows from (S.53), as well as (S.54) and (S.55) then leads to

$$\sup_{t\in\mathbb{R}}|\hat{F}_{iK}(t)-\tilde{F}_{iK}(t)|=O_p(\alpha_n). \tag{S.57}$$

The result in (4.18) then follows from (S.52), (S.57) and the triangle inequality.

For the next result in (4.19), similarly as before we first start by showing that $\|\tilde{f}_{iK}-f_{iK}\|_{L^2(\mathbb{R})}=O_p(a_n+b_n)$, where $\tilde{f}_i(t):=\tilde{F}_i'(t)=\varphi((t-\tilde{u}_{in})/\tilde{\sigma}_{in})/\tilde{\sigma}_{in}$. Since $f_i(t)=F_i'(t)=\varphi((t-u_i)/\sigma_i)/\sigma_i$, we have

$$\begin{aligned}\Big\|\frac{1}{\tilde{\sigma}_{in}}\varphi\left(\frac{\cdot-\tilde{u}_{in}}{\tilde{\sigma}_{in}}\right)-\frac{1}{\sigma_i}\varphi\left(\frac{\cdot-u_i}{\sigma_i}\right)\Big\|_{L^2(\mathbb{R})}&\le\frac{1}{\tilde{\sigma}_{in}}\Big\|\varphi\left(\frac{\cdot-\tilde{u}_{in}}{\tilde{\sigma}_{in}}\right)-\varphi\left(\frac{\cdot-u_i}{\sigma_i}\right)\Big\|_{L^2(\mathbb{R})}\\&\quad+\Big|\frac{1}{\tilde{\sigma}_{in}}-\frac{1}{\sigma_i}\Big|\,\Big\|\varphi\left(\frac{\cdot-u_i}{\sigma_i}\right)\Big\|_{L^2(\mathbb{R})}.\end{aligned} \tag{S.58}$$

Thus, since $\|\varphi\left(\frac{\cdot-u_i}{\sigma_i}\right)\|_{L^2(\mathbb{R})}=O(\sigma_i^{1/2})$ and $|\tilde{\sigma}_{in}^{-1}-\sigma_i^{-1}|=O_p(a_n+b_n)$, we obtain

$$\Big|\frac{1}{\tilde{\sigma}_{in}}-\frac{1}{\sigma_i}\Big|\,\Big\|\varphi\left(\frac{\cdot-u_i}{\sigma_i}\right)\Big\|_{L^2(\mathbb{R})}=O_p(a_n+b_n). \tag{S.59}$$

Using the relation $\varphi'(t)=-t\varphi(t)$ and a Taylor expansion, it follows that

$$\Big\|\varphi\left(\frac{\cdot-\tilde{u}_{in}}{\tilde{\sigma}_{in}}\right)-\varphi\left(\frac{\cdot-u_i}{\sigma_i}\right)\Big\|^2_{L^2(\mathbb{R})}=\int_{\mathbb{R}}(\varphi'(\varepsilon_t))^2\Delta_{in}^2(t)dt=\int_{\mathbb{R}}\varepsilon_t^2\varphi^2(\varepsilon_t)\Delta_{in}^2(t)dt,$$

where ε_t is between $r_{in}(t)$ and $r_i(t)$. Hence, from (S.58) and (S.59) it suffices to show that $\int_{\mathbb{R}}\varepsilon_t^2\varphi^2(\varepsilon_t)\Delta_{in}^2(t)dt=O_p((a_n+b_n)^2)$. Indeed, from the fact that $|\varepsilon_t|\le|r_{in}(t)|+|r_i(t)|$, $\sup_{t\in I_{in}}|r_{in}(t)|=O_p(a_n+b_n)$, $\sup_{t\in I_{in}}|r_i(t)|=O_p(a_n+$

b_n) and $\varphi(\varepsilon_t)1_{\{t\in I_{in}^c\}} \le \varphi(r_{in}(t)) + \varphi(r_i(t))$, we have

$$\begin{aligned}
&\int_{\mathbb{R}} \varepsilon_t^2\varphi^2(\varepsilon_t)\Delta_{in}^2(t)dt \\
&= \int_{I_{in}} \varepsilon_t^2\varphi^2(\varepsilon_t)\Delta_{in}^2(t)dt + \int_{I_{in}^c} \varepsilon_t^2\varphi^2(\varepsilon_t)\Delta_{in}^2(t)dt \\
&\le \varphi^2(0)O_p((a_n+b_n)^5) + \int_{I_{in}^c} [\varphi(r_{in}(t)) + \varphi(r_i(t))]^2 (r_{in}(t) + r_i(t))^2\Delta_{in}^2(t)dt \\
&\le O_p((a_n+b_n)^5) + \int_{\mathbb{R}} [\varphi(r_{in}(t)) + \varphi(r_i(t))]^2 (r_{in}(t) + r_i(t))^2\Delta_{in}^2(t)dt,
\end{aligned} \tag{S.60}$$

where the first inequality follows from (S.50) and the relation $|r_{in}(t)||r_i(t)|1_{\{t\in I_{in}^c\}} = r_{in}(t)r_i(t)1_{\{t\in I_{in}^c\}}$. From $\int_{\mathbb{R}} \varphi^2(s)|s|^p ds < \infty$, $p \in \mathbb{N}$, we obtain the following facts:

$$\begin{aligned}
\int_{\mathbb{R}} \varphi^2(r_{in}(t))r_{in}^2(t)\Delta_{in}^2(t)dt &\le \sigma_i^{-2}O_p((a_n+b_n)^2), \\
\int_{\mathbb{R}} \varphi^2(r_{in}(t))r_i^2(t)\Delta_{in}^2(t)dt &\le \sigma_i^{-4}O_p((a_n+b_n)^2), \\
\Big|\int_{\mathbb{R}} \varphi^2(r_{in}(t))r_{in}(t)r_i(t)\Delta_{in}^2(t)dt\Big| &\le \sigma_i^{-3}O_p((a_n+b_n)^2), \\
\Big|\int_{\mathbb{R}} \varphi^2(r_i(t))r_{in}(t)r_i(t)\Delta_{in}^2(t)dt\Big| &\le (\boldsymbol{\beta}_K^T\boldsymbol{\beta}_K\lambda_{\min}(\boldsymbol{\Sigma}_{iK}))^{-3/2}O_p((a_n+b_n)^2), \\
\int_{\mathbb{R}} \varphi^2(r_i(t))r_i^2(t)\Delta_{in}^2(t)dt &\le (\boldsymbol{\beta}_K^T\boldsymbol{\beta}_K\lambda_{\min}(\boldsymbol{\Sigma}_{iK}))^{-1}O_p((a_n+b_n)^2), \\
\int_{\mathbb{R}} \varphi^2(r_i(t))r_{in}^2(t)\Delta_{in}^2(t)dt &\le (\boldsymbol{\beta}_K^T\boldsymbol{\beta}_K\lambda_{\min}(\boldsymbol{\Sigma}_{iK}))^{-2}O_p((a_n+b_n)^2), \\
\Big|\int_{\mathbb{R}} \varphi(r_{in}(t))\varphi(r_i(t))r_i(t)r_{in}(t)\Delta_{in}^2(t)dt\Big| &\le \sigma_i^{-3}O_p((a_n+b_n)^2).
\end{aligned}$$

These facts along with (S.60) imply $\int_{\mathbb{R}} \varepsilon_t^2\varphi^2(\varepsilon_t)\Delta_{in}^2(t)dt \le O_p((a_n+b_n)^5) + O_p((a_n+b_n)^2) = O_p((a_n+b_n)^2)$ and

$$\left\|\tilde{f}_{iK} - f_{iK}\right\|_{L^2(\mathbb{R})} = O_p(a_n+b_n).$$

Similar arguments imply $\|\hat{f}_{iK} - \tilde{f}_{iK}\|_{L^2(\mathbb{R})} = O_p(\alpha_n)$ and the result in (4.19).

Finally, from condition (C1) we have $\lambda_{\min}(\boldsymbol{\Sigma}_{iK}) \geq \kappa_0$ and also $\sigma_i^2 = (\boldsymbol{\beta}_K^T \boldsymbol{\Sigma}_{iK} \boldsymbol{\beta}_K) \geq \boldsymbol{\beta}_K^T \boldsymbol{\beta}_K \lambda_{\min}(\boldsymbol{\Sigma}_{iK}) \geq \boldsymbol{\beta}_K^T \boldsymbol{\beta}_K \kappa_0$ a.s., which implies $\sigma_i^{-1} = O(1)$ and $\lambda_{\min}(\boldsymbol{\Sigma}_{iK})^{-1} = O(1)$ a.s., where the $O(1)$ terms are uniform in i. Since $\left\|\boldsymbol{\Sigma}_{iK} - \hat{\boldsymbol{\Sigma}}_{iK}\right\|_F = O(a_n + b_n)$ a.s. as $n \to \infty$, where the $O(a_n + b_n)$ term is uniform over i, and $\left\|\hat{\boldsymbol{\xi}}_{iK} - \tilde{\boldsymbol{\xi}}_{iK}\right\|_2 = O_p(a_n + b_n)$, where the $O_p(a_n + b_n)$ term is also uniform over i, it can be easily checked from the previous arguments that the rates of convergence in (4.17), (4.18) and (4.19) are uniform in i. □

Proof of Theorem 6. Recall that $\eta_{iK} := \beta_0 + \boldsymbol{\beta}_K^T \boldsymbol{\xi}_{iK}$ is the K-truncated linear predictor for the ith subject and $\tilde{\eta}_{iK} := \beta_0 + \boldsymbol{\beta}_K^T \tilde{\boldsymbol{\xi}}_{iK}$ its best prediction. Also, recall that $\mathcal{P}_{iK}$ corresponds to the predictive distribution of η_{iK} given $\mathbf{X}_i$ and $\mathbf{T}_i$, and $\hat{\mathcal{P}}_{iK}$ is the corresponding estimate. Writing $Y_i = \beta_0 + \boldsymbol{\beta}_K^T \boldsymbol{\xi}_{iK} + \sum_{k \geq K+1} \beta_k \xi_{ik} + \epsilon_{iY} = \eta_{iK} + R_{iK} + \epsilon_{iY}$, where $R_{iK} = \sum_{k \geq K+1} \beta_k \xi_{ik}$, the estimated Wasserstein discrepancy is given by $\hat{\mathcal{D}}_{nK} = n^{-1} \sum_{i=1}^n W_2^2(\delta_{Y_i}, \hat{\mathcal{P}}_{iK})$, where

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^n \mathcal{W}_2^2(\mathcal{A}_{Y_i}, \hat{\mathcal{P}}_{iK}) \\
& = n^{-1}\sum_{i=1}^n (Y_i - \hat{\eta}_{iK})^2 + n^{-1}\sum_{i=1}^n \hat{\boldsymbol{\beta}}_K^T \hat{\boldsymbol{\Sigma}}_{iK} \hat{\boldsymbol{\beta}}_K \\
& = n^{-1}\sum_{i=1}^n (\eta_{iK} - \hat{\eta}_{iK})^2 + n^{-1}\sum_{i=1}^n \epsilon_{iY}^2 + n^{-1}\sum_{i=1}^n R_{iK}^2 + 2n^{-1}\sum_{i=1}^n (\eta_{iK} - \hat{\eta}_{iK})\epsilon_{iY} \\
& \quad + 2n^{-1}\sum_{i=1}^n (\eta_{iK} - \hat{\eta}_{iK}) R_{iK} + 2n^{-1}\sum_{i=1}^n R_{iK}\epsilon_{iY} + n^{-1}\sum_{i=1}^n \hat{\boldsymbol{\beta}}_K^T \hat{\boldsymbol{\Sigma}}_{iK} \hat{\boldsymbol{\beta}}_K.
\end{aligned}
\tag{S.61}
$$

Since $n_i = m_0 < N_0$, by the central limit theorem,

$$
n^{-1}\sum_{i=1}^n (\eta_{iK} - \tilde{\eta}_{iK}) R_{iK} = -\boldsymbol{\beta}_K^T E\left(\boldsymbol{\Lambda}_K \boldsymbol{\Phi}_{1K}^T \boldsymbol{\Sigma}_1^{-1} \sum_{k \geq K+1} \phi_k(\mathbf{T}_1)\lambda_k\beta_k\right) + O_p(n^{-1/2}),
$$

and

$$
n^{-1}\sum_{i=1}^n R_{iK}^2 = \sum_{k \geq K+1} \beta_k^2 \lambda_k + O_p(n^{-1/2}). \tag{S.62}
$$

Combining this with $n^{-1}\sum_{i=1}^n (\tilde{\eta}_{iK} - \hat{\eta}_{iK})^2 = O_p(\alpha_n^2)$, as shown in the proof of Lemma S13,

$$
\begin{aligned}
n^{-1}\sum_{i=1}^n (\eta_{iK} - \hat{\eta}_{iK}) R_{iK} &= n^{-1}\sum_{i=1}^n (\eta_{iK} - \tilde{\eta}_{iK}) R_{iK} + n^{-1}\sum_{i=1}^n (\tilde{\eta}_{iK} - \hat{\eta}_{iK}) R_{iK} \\
&= -\boldsymbol{\beta}_K^T E\left(\boldsymbol{\Lambda}_K \boldsymbol{\Phi}_{1K}^T \boldsymbol{\Sigma}_1^{-1} \sum_{k \geq K+1} \phi_k(\mathbf{T}_1)\lambda_k\beta_k\right) + O_p(\alpha_n).
\end{aligned}
\tag{S.63}
$$

Next

$$
\begin{aligned}
n^{-1}\sum_{i=1}^n (\eta_{iK} - \hat{\eta}_{iK})\epsilon_{iY} &= n^{-1}\sum_{i=1}^n (\eta_{iK} - \tilde{\eta}_{iK})\epsilon_{iY} + n^{-1}\sum_{i=1}^n (\tilde{\eta}_{iK} - \hat{\eta}_{iK})\epsilon_{iY} \\
&= O_p(n^{-1/2}) + O_p(\alpha_n) = O_p(\alpha_n),
\end{aligned}
\tag{S.64}
$$

where the last equality follows from Lemma S13 and since $n^{-1}\sum_{i=1}^n(\eta_{iK} - \tilde{\eta}_{iK})\epsilon_{iY} = O_p(n^{-1/2})$, which is due to the Central Limit Theorem. Similarly, from Lemma S14 we have

$$n^{-1}\sum_{i=1}^n(\eta_{iK} - \tilde{\eta}_{iK})^2 = \boldsymbol{\beta}_K^T E(\boldsymbol{\Sigma}_{1K})\boldsymbol{\beta}_K + O_p(n^{-1/2}), \tag{S.65}$$

and

$$\begin{aligned} n^{-1}\sum_{i=1}^n(\eta_{iK} - \hat{\eta}_{iK})^2 - n^{-1}\sum_{i=1}^n(\eta_{iK} - \tilde{\eta}_{iK})^2 &= n^{-1}\sum_{i=1}^n(\tilde{\eta}_{iK} - \hat{\eta}_{iK})^2 \\ &+ 2n^{-1}\sum_{i=1}^n(\eta_{iK} - \tilde{\eta}_{iK})(\tilde{\eta}_{iK} - \hat{\eta}_{iK} = O_p(\alpha_n), \end{aligned} \tag{S.66}$$

where the last equality follows from the fact that $n^{-1}\sum_{i=1}^n(\tilde{\eta}_{iK} - \hat{\eta}_{iK})^2 = O_p(\alpha_n^2)$, (S.65) and the Cauchy–Schwarz inequality. Combining (S.65) and (S.66) leads to

$$n^{-1}\sum_{i=1}^n(\eta_{iK} - \hat{\eta}_{iK})^2 = \boldsymbol{\beta}_K^T E(\boldsymbol{\Sigma}_{1K})\boldsymbol{\beta}_K + O_p(\alpha_n). \tag{S.67}$$

We further note that

$$\begin{aligned} &|\hat{\boldsymbol{\beta}}_K^T\hat{\boldsymbol{\Sigma}}_{iK}\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K^T\boldsymbol{\Sigma}_{iK}\boldsymbol{\beta}_K| \\ &= |\hat{\boldsymbol{\beta}}_K^T(\hat{\boldsymbol{\Sigma}}_{iK} - \boldsymbol{\Sigma}_{iK})\hat{\boldsymbol{\beta}}_K + (\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K)^T\boldsymbol{\Sigma}_{iK}\hat{\boldsymbol{\beta}}_K + \boldsymbol{\beta}_K^T\boldsymbol{\Sigma}_{iK}(\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K)| \\ &\leq \|\hat{\boldsymbol{\beta}}_K\|_2^2\|\hat{\boldsymbol{\Sigma}}_{iK} - \boldsymbol{\Sigma}_{iK}\|_{\text{op},2} + \|\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K\|_2\|\boldsymbol{\Sigma}_{iK}\|_{\text{op},2}(\|\hat{\boldsymbol{\beta}}_K\|_2 + \|\boldsymbol{\beta}_K\|_2). \end{aligned}$$

From the proof of Theorem 5, we have $\left\|\boldsymbol{\Sigma}_{iK} - \hat{\boldsymbol{\Sigma}}_{iK}\right\|_F = O(a_n + b_n)$ a.s. as $n \to \infty$, where the $O(a_n + b_n)$ term is uniform in i. Since $\|\boldsymbol{\Sigma}_{iK}\|_F = O(1)$

uniformly over i,

$$\left|n^{-1}\sum_{i=1}^{n}(\hat{\boldsymbol{\beta}}_K^T\hat{\boldsymbol{\Sigma}}_{iK}\hat{\boldsymbol{\beta}}_K-\boldsymbol{\beta}_K^T\boldsymbol{\Sigma}_{iK}\boldsymbol{\beta}_K)\right| \le n^{-1}\sum_{i=1}^{n}|\hat{\boldsymbol{\beta}}_K^T\hat{\boldsymbol{\Sigma}}_{iK}\hat{\boldsymbol{\beta}}_K-\boldsymbol{\beta}_K^T\boldsymbol{\Sigma}_{iK}\boldsymbol{\beta}_K|$$

$$\le \|\hat{\boldsymbol{\beta}}_K-\boldsymbol{\beta}_K\|_2(\|\hat{\boldsymbol{\beta}}_K\|_2+\|\boldsymbol{\beta}_K\|_2)\, n^{-1}\sum_{i=1}^{n}\|\boldsymbol{\Sigma}_{iK}\|_F$$

$$+\|\hat{\boldsymbol{\beta}}_K\|_2^2\, n^{-1}\sum_{i=1}^{n}\|\hat{\boldsymbol{\Sigma}}_{iK}-\boldsymbol{\Sigma}_{iK}\|_F$$

$$\le \|\hat{\boldsymbol{\beta}}_K-\boldsymbol{\beta}_K\|_2(\|\hat{\boldsymbol{\beta}}_K\|_2+\|\boldsymbol{\beta}_K\|_2)O(1)$$

$$+\|\hat{\boldsymbol{\beta}}_K\|_2^2 O(a_n+b_n) \quad \text{a.s.},$$

as $n\to\infty$. From Lemma S11, we have $\|\hat{\boldsymbol{\beta}}_K-\boldsymbol{\beta}_K\|_2=O_p(\alpha_n)$, which combined with $\|\hat{\boldsymbol{\beta}}_K\|_2\le\|\hat{\boldsymbol{\beta}}_K-\boldsymbol{\beta}_K\|_2+\|\boldsymbol{\beta}_K\|_2=O_p(1)$ leads to

$$n^{-1}\sum_{i=1}^{n}\hat{\boldsymbol{\beta}}_K^T\hat{\boldsymbol{\Sigma}}_{iK}\hat{\boldsymbol{\beta}}_K-n^{-1}\sum_{i=1}^{n}\boldsymbol{\beta}_K^T\boldsymbol{\Sigma}_{iK}\boldsymbol{\beta}_K=O_p(\alpha_n).$$

This along with an application of the Central Limit Theorem shows that

$$n^{-1}\sum_{i=1}^{n}\hat{\boldsymbol{\beta}}_K^T\hat{\boldsymbol{\Sigma}}_{iK}\hat{\boldsymbol{\beta}}_K=\boldsymbol{\beta}_K^T E(\boldsymbol{\Sigma}_{1K})\boldsymbol{\beta}_K+O_p(\alpha_n). \tag{S.68}$$

Finally, it is easy to show that $n^{-1}\sum_{i=1}^{n}R_{iK}\epsilon_{iY}=O_p(n^{-1/2})$ and $n^{-1}\sum_{i=1}^{n}\epsilon_{iY}^2=\sigma_Y^2+O_p(n^{-1/2})$, applying the CLT. Combining with (S.62), (S.63), (S.64), (S.67), and (S.68),

$$\hat{\mathcal{D}}_{nK}=2\boldsymbol{\beta}_K^T E(\boldsymbol{\Sigma}_{1K})\boldsymbol{\beta}_K+\sigma_Y^2+\sum_{k\ge K+1}\beta_k^2\lambda_k-2\boldsymbol{\beta}_K^T E\left(\boldsymbol{\Lambda}_K\boldsymbol{\Phi}_{1K}^T\boldsymbol{\Sigma}_1^{-1}\sum_{k\ge K+1}\phi_k(\mathbf{T}_1)\lambda_k\beta_k\right)$$

$$+O_p(\alpha_n),$$

implying the first result in (4.20). Similar arguments show that the Wasserstein distance using true population quantities $\mathcal{D}_{nK}$ is such that

$$\mathcal{D}_{nK} = n^{-1}\sum_{i=1}^{n}\mathcal{W}_2^2(\mathcal{A}_{Y_i},\mathcal{P}_{iK}) = n^{-1}\sum_{i=1}^{n}(Y_i-\tilde{\eta}_{iK})^2 + n^{-1}\sum_{i=1}^{n}\boldsymbol{\beta}_K^T\boldsymbol{\Sigma}_{iK}\boldsymbol{\beta}_K$$
$$= \mathcal{D}_K + O_p(n^{-1/2}),$$

where

$$\mathcal{D}_K = 2\boldsymbol{\beta}_K^T E(\boldsymbol{\Sigma}_{1K})\boldsymbol{\beta}_K + \sigma_Y^2 + \sum_{k\geq K+1}\beta_k^2\lambda_k$$
$$- 2\boldsymbol{\beta}_K^T E\left(\boldsymbol{\Lambda}_K\boldsymbol{\Phi}_{1K}^T\boldsymbol{\Sigma}_1^{-1}\sum_{k\geq K+1}\phi_k(\mathbf{T}_1)\lambda_k\beta_k\right).$$

Since $Y = \mu_Y + \int_{\mathcal{T}}\beta(t)U(t) + \epsilon_Y$, where $\mu_Y = E(Y)$ and $U(t) = X(t) - \mu(t)$, we have $E(Y^2) = \mu_Y^2 + \sigma_Y^2 + E(\langle\beta,U\rangle_{L^2}^2)$, where $\langle\cdot,\cdot\rangle_L^2$ is the $L^2(\mathcal{T})$ inner product. From (X4) it follows that $E(\langle\beta,U\rangle_{L^2}^2) = \sum_{j=1}^{\infty}\beta_j^2\lambda_j$ as the FPCs are independent in the Gaussian case. Then

$$n^{-1}\sum_{i=1}^{n}(Y_i-\bar{Y}_n)^2 = \text{Var}(Y) + O_p(n^{-1/2}) = \sigma_Y^2 + \sum_{j=1}^{\infty}\lambda_j\beta_j^2 + O_p(n^{-1/2}). \tag{S.69}$$

Also, $|\hat{\beta}_j| \leq \|\hat{\beta}_M\|_{L^2}$ and $|\beta_j| \leq \|\beta\|_{L^2}$. With perturbation results as used in the proof of Lemma S10 this leads to

$$\Big|\sum_{m=1}^{M}\hat{\lambda}_m\hat{\beta}_m^2 - \lambda_m\beta_m^2\Big|$$
$$\leq \sum_{m=1}^{M}|\hat{\lambda}_m-\lambda_m||\hat{\beta}_m^2-\beta_m^2| + \sum_{m=1}^{M}|\hat{\lambda}_m-\lambda_m|\beta_m^2 + \sum_{m=1}^{M}\lambda_m|\hat{\beta}_m^2-\beta_m^2|$$
$$\leq \|\hat{\Xi}-\Xi\|_{\text{op}}(\|\hat{\beta}_M\|_{L^2}+\|\beta\|_{L^2})\sum_{m=1}^{M}|\hat{\beta}_m-\beta_m| + \|\hat{\Xi}-\Xi\|_{\text{op}}\sum_{m=1}^{M}\beta_m^2$$
$$+ (\|\hat{\beta}_M\|_{L^2}+\|\beta\|_{L^2})\sum_{m=1}^{M}\lambda_m|\hat{\beta}_m-\beta_m|. \tag{S.70}$$

From the proof of Lemma S11 and since $\sum_{j=1}^{\infty} \lambda_j < \infty$, we have

$$
\begin{aligned}
&\sum_{m=1}^{M} \lambda_m |\hat{\beta}_m - \beta_m| \\
&\leq \|\hat{\beta}_M - \beta\|_{L^2} \sum_{m=1}^{M} \lambda_m \|\hat{\phi}_m - \phi_m\|_{L^2} + \|\hat{\beta}_M - \beta\|_{L^2} \left(\sum_{m=1}^{M} \lambda_m \right) \\
&\quad + \|\beta\|_{L^2} \sum_{m=1}^{M} \lambda_m \|\hat{\phi}_m - \phi_m\|_{L^2} \\
&\leq \left(\sum_{j=1}^{\infty} \lambda_j \right) \|\hat{\beta}_M - \beta\|_{L^2} + 2\sqrt{2} \|\hat{\Xi} - \Xi\|_{\text{op}} \left(\|\hat{\beta}_M - \beta\|_{L^2} + \|\beta\|_{L^2} \right) \left(\sum_{m=1}^{M} \frac{\lambda_m}{\delta_m} \right) \quad \text{a.s.} \\
&\leq O_p(\alpha_n) + O_p(1) O(c_n^{\rho}) = O_p(\alpha_n), \qquad \text{(S.71)}
\end{aligned}
$$

where the last inequality follows from Lemma S15 and Lemma S11. Similarly

$$
\begin{aligned}
&\sum_{m=1}^{M} |\hat{\beta}_m - \beta_m| \\
&\leq 2\sqrt{2} \|\hat{\Xi} - \Xi\|_{\text{op}} \left(\|\hat{\beta}_M - \beta\|_{L^2} + \|\beta\|_{L^2} \right) \left(\sum_{m=1}^{M} \frac{1}{\delta_m} \right) + \|\hat{\beta}_M - \beta\|_{L^2} M \quad \text{a.s.} \\
&\leq O(c_n) O(c_n^{\rho-1}) \left(\|\hat{\beta}_M - \beta\|_{L^2} + \|\beta\|_{L^2} \right) + \|\hat{\beta}_M - \beta\|_{L^2} O(c_n^{\rho-1}) \quad \text{a.s.} \\
&\leq O_p(c_n^{\rho}) + O_p(c_n^{\rho-1} \alpha_n), \qquad \text{(S.72)}
\end{aligned}
$$

where the second and third inequalities follow from Lemma S11 and using that $\sum_{m=1}^{M} \delta_m^{-1} = O(c_n^{\rho-1})$, which was shown in the proof of Lemma S10, along with the fact that $M = O(c_n^{\rho-1})$, which is due to the condition $\sum_{m=1}^{M} \frac{1}{\sqrt{\lambda_m}\delta_m} = O(c_n^{\rho-1})$ and $0 < \delta_m < \lambda_m \leq \lambda_1$. Combining (S.70), (S.71) and (S.72) leads to

$$
\Big| \sum_{m=1}^{M} \hat{\lambda}_m \hat{\beta}_m^2 - \sum_{m=1}^{M} \lambda_m \beta_m^2 \Big| = O_p(\alpha_n).
$$

This implies

$$\Big|\sum_{m=1}^{M}\hat{\lambda}_m\hat{\beta}_m^2-\sum_{m=1}^{\infty}\lambda_m\beta_m^2\Big|\leq O_p(\alpha_n)+\sum_{m\geq M+1}\lambda_m\beta_m^2,$$

and the result in (4.21) follows from (S.69). □

Proof of Theorem 7. Note that

$$\begin{aligned}\|\tilde{\boldsymbol{\xi}}_K^*-\boldsymbol{\xi}_K^*\|_2^2&=\sum_{k=1}^{K}(\lambda_k\boldsymbol{\phi}_k^{*T}\boldsymbol{\Sigma}^{*-1}(\mathbf{X}^*-\boldsymbol{\mu}^*)-\xi_k^*)^2\\&\lesssim\sum_{k=1}^{K}(\mathbf{e}_k^{*T}\boldsymbol{\Sigma}^{*-1}(\mathbf{X}^*-\boldsymbol{\mu}^*))^2+\sum_{k=1}^{K}(\boldsymbol{\phi}_k^{*T}\mathbf{W}^*(\mathbf{Y}^*-\boldsymbol{\mu}^*)-\xi_k^*)^2\\&+\sum_{k=1}^{K}(\boldsymbol{\phi}_k^{*T}\mathbf{W}^*\boldsymbol{\epsilon}^*)^2.\end{aligned}$$

Similar to the proof of Theorem 3, we have

$$|\boldsymbol{\phi}_k^{*T}\mathbf{W}^*(\mathbf{Y}^*-\boldsymbol{\mu}^*)-\xi_k^*|\leq\lambda_k^{-1}\left(\sum_{l=1}^{m}w_l^{*2}+(1-T^{*(m)})^2+(1-T^{*(m)})\right),$$

where $T^{*(m)}=\max_{j=1,\ldots,m^*}\mathbf{T}_j^*$. This implies

$$E\left(\sum_{k=1}^{K}(\boldsymbol{\phi}_k^{*T}\mathbf{W}^*(\mathbf{Y}^*-\boldsymbol{\mu}^*)-\xi_k^*)^2\right)=O(m^{*-2}).$$

Also

$$E\left(\sum_{k=1}^{K}(\boldsymbol{\phi}_k^{*T}\mathbf{W}^*\boldsymbol{\epsilon}^*)^2\right)=O(m^{*-1}),$$

and

$$E\left((\mathbf{e}_k^{*T}\boldsymbol{\Sigma}^{*-1}(\mathbf{X}^*-\boldsymbol{\mu}^*))^2\right)=O(m^{*-1}).$$

Therefore

$$E\left(\|\tilde{\boldsymbol{\xi}}_K^* - \boldsymbol{\xi}_K^*\|_2^2\right) = O(m^{*-1}). \tag{S.73}$$

Recall that $\mathcal{P}_K^* \overset{d}{=} N(\beta_0+\boldsymbol{\beta}_K^T\tilde{\boldsymbol{\xi}}_K^*, \boldsymbol{\beta}_K^T\boldsymbol{\Sigma}_K^*\boldsymbol{\beta}_K)$. By construction of the 2-Wasserstein distance,

$$\begin{aligned}\mathcal{W}_2^2(\mathcal{P}_K^*, \mathcal{A}_{\beta_0+\boldsymbol{\beta}_K^T\boldsymbol{\xi}_K^*}) &= (\boldsymbol{\beta}_K^T(\tilde{\boldsymbol{\xi}}_K^* - \boldsymbol{\xi}_K^*))^2 + \boldsymbol{\beta}_K^T\boldsymbol{\Sigma}_K^*\boldsymbol{\beta}_K \\ &\leq \|\boldsymbol{\beta}_K\|_2^2\|\tilde{\boldsymbol{\xi}}_K^* - \boldsymbol{\xi}_K^*\|_2^2 + \|\boldsymbol{\beta}_K\|_2^2\|\boldsymbol{\Sigma}_K^*\|_{\text{op},2} \\ &= O_p(m^{*-1}),\end{aligned}$$

where the last equality is due to (S.73) and using that $\|\boldsymbol{\Sigma}_K^*\|_{\text{op},2} = O_p(m^{*-1})$, which follows analogously as in the proof of Theorem 2. This shows the first result. Next,

$$\begin{aligned}\mathcal{W}_2^2(\hat{\mathcal{P}}_K^*, \mathcal{A}_{\beta_0+\boldsymbol{\beta}_K^T\boldsymbol{\xi}_K^*}) &= \hat{\boldsymbol{\beta}}_K^T\hat{\boldsymbol{\Sigma}}_K^*\hat{\boldsymbol{\beta}}_K + (\hat{\beta}_0 + \hat{\boldsymbol{\beta}}_K^T\hat{\boldsymbol{\xi}}_K^* - \beta_0 - \boldsymbol{\beta}_K^T\boldsymbol{\xi}_K^*)^2 \\ &\lesssim \|\hat{\boldsymbol{\beta}}_K\|_2^2\|\hat{\boldsymbol{\Sigma}}_K^* - \boldsymbol{\Sigma}_K^*\|_{\text{op},2} + \|\hat{\boldsymbol{\beta}}_K\|_2^2\|\boldsymbol{\Sigma}_K^*\|_{\text{op},2} + (\hat{\beta}_0 - \beta_0)^2 \\ &+ \|\hat{\boldsymbol{\beta}}_K\|_2^2\|\hat{\boldsymbol{\xi}}_K^* - \boldsymbol{\xi}_K^*\|_2^2 + \|\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K\|_2^2\|\boldsymbol{\xi}_K^*\|_2^2 \\ &= O_p\left(m^{*2}(a_n + b_n)^2 + m^{*-1} + a_n + b_n + r_n^{*2}\right), \end{aligned} \tag{S.74}$$

where the last equality is due to Theorem 1, Theorem 2, the fact that $\|\boldsymbol{\xi}_K^*\|_2 = O_p(1)$, $\|\boldsymbol{\Sigma}_K^*\|_{\text{op},2} = O_p(m^{*-1})$, and using Lemma S11 with $h = n^{-1/3}$. The second result follows. □

In the following, we say that a process X is explained by its first K principal components if $X(t) = \mu(t) + \sum_{k=1}^K \xi_k\phi_k(t)$ and thus is of finite dimension K.

Lemma S1. *Suppose that the process X is finite dimensional and explained by its first $K = 2$ principal components. If ϕ_1 and ϕ_2 are bijective and differentiable in a finite partition of $\mathcal{T}$, then $\mathbf{\Sigma}_{iK}$ has a positive eigengap almost surely.*

Proof of Lemma S1. Recalling that $\mathbf{\Sigma}_{iK} = \mathbf{\Lambda}_K - \mathbf{\Lambda}_K \mathbf{\Phi}_{iK}^T \mathbf{\Sigma}_i^{-1} \mathbf{\Phi}_{iK} \mathbf{\Lambda}_K$ and since $K = 2$, it follows that the characteristic polynomial of $\mathbf{\Sigma}_{iK}$ is given by $p(\lambda) = \lambda^2 - \operatorname{tr}(\mathbf{\Sigma}_{iK})\lambda + \det(\mathbf{\Sigma}_{iK})$, and thus the eigengap is equal to $\sqrt{\Delta_p}$, where Δ_p is the discriminant of the quadratic polynomial p. It is easy to show that

$$\Delta_p = (\lambda_1 - \lambda_2 + \lambda_2^2 \phi_{i2}^T \mathbf{\Sigma}_i^{-1} \phi_{i2} - \lambda_1^2 \phi_{i1}^T \mathbf{\Sigma}_i^{-1} \phi_{i1})^2 + 4\lambda_1^2 \lambda_2^2 (\phi_{i1}^T \mathbf{\Sigma}_i^{-1} \phi_{i2})^2,$$

so that it suffices to check that $\phi_{i1}^T \mathbf{\Sigma}_i^{-1} \phi_{i2}$ is not identically zero almost surely. Let $B = \sigma^2 I_{n_i} + \lambda_1 \phi_{i1} \phi_{i1}^T$, where I_{n_i} denotes the $n_i \times n_i$ identity matrix, and denote by $\|\cdot\|_2$ the Euclidean norm in $\mathbb{R}^{n_i}$. By the Sherman-Morrison formula, it follows that $B^{-1} = \sigma^{-2} \left(I_{n_i} - \frac{\lambda_1 \phi_{i1} \phi_{i1}^T}{\sigma^2 + \lambda_1 \|\phi_{i1}\|_2^2} \right)$, and a second application of the formula leads to

$$\mathbf{\Sigma}_i^{-1} = B^{-1} - \frac{B^{-1} \lambda_2 \phi_{i2} \phi_{i2}^T B^{-1}}{1 + \lambda_2 \phi_{i2}^T B^{-1} \phi_{i2}}.$$

Thus

$$\phi_{i1}^T \mathbf{\Sigma}_i^{-1} \phi_{i2} = \frac{\phi_{i1}^T B^{-1} \phi_{i2}}{1 + \lambda_2 \phi_{i2}^T B^{-1} \phi_{i2}},$$

where $\phi_{i1}^T B^{-1} \phi_{i2} = \frac{\phi_{i1}^T \phi_{i2}}{\sigma^2 + \lambda_1 \|\phi_{i1}\|_2^2}$ and $\phi_{i2}^T B^{-1} \phi_{i2} > 0$ a.s. since the eigenvalues of B are bounded below by σ^2. The conclusion then follows if we can show that $\phi_{i1}^T \phi_{i2} \neq 0$ almost surely. Note that $\phi_{i1}^T \phi_{i2} = \sum_{j=1}^{n_i} \phi_1(T_{ij}) \phi_2(T_{ij})$ and the T_{ij} are i.i.d. with a continuous distribution supported on $\mathcal{T}$. Thus, the distribution of $\phi_{i1}^T \phi_{i2}$ corresponds to the n-fold convolution of the continuous distribution associated with $\phi_1(T_{i1})\phi_2(T_{i1})$, which is a continuous probability measure, and hence $\phi_{i1}^T\phi_{i2} \neq 0$ holds almost surely. $\square$

Lemma S2. *Let $T_1, \dots, T_m$ be i.i.d. with density function $f(t)$, $t \in \mathcal{T} = [0,1]$ and let $T_{(1)}, \dots, T_{(m)}$ be the order statistics. Let $w_l := T_{(l)} - T_{(l-1)}$, $l = 1, \dots, m$, where $T_{(0)} := 0$, be the spacing between the order statistics. Suppose that there exists $c_0 > 0$ such that $f(t) \geq c_0$ for all $t \in \mathcal{T}$. Then, for any integer $p \geq 1$ it holds that,*

$$E(w_l^p) = O(m^{-p}), \quad l = 1, \dots, m,$$

and

$$E[\big(1 - T_{(m)}\big)^p] = O(m^{-p}).$$

Proof of Lemma S2. One can replace T_l with i.i.d. copies $Q(U_l)$, $l = 1, \dots, m$, where the $U_l \overset{iid}{\sim} U(0,1)$ and Q is the quantile function corresponding to f. Since f is strictly positive, then $T_{(l)} = Q(U_{(l)})$, $l = 1, \dots, m$. From a Taylor expansion of $Q(\cdot)$, we have

$$E\left(w_l^p\right) = E[Q'(\eta_l)(U_{(l)} - U_{(l-1)})]^p \leq c_0^{-p} E[U_{(l)} - U_{(l-1)}]^p,$$

where η_l is between $U_{(l-1)}$ and $U_{(l)}$, and the last inequality follows from the fact that $Q'(t) = 1/f(Q(t)) \leq c_0^{-1}$. The first result follows since $U_{(l)} - U_{(l-1)} \sim \text{Beta}(1, m)$ which implies $E[U_{(l)} - U_{(l-1)}]^p = O(m^{-p})$. Similarly, by expanding $Q(U_{(m)})$ around $Q(1) = 1$ and since it can be verified that $E[(1 - U_{(m)})^p] = m!p!/(m+p)! = O(m^{-p})$, the second result follows. $\square$

The next two lemmas are for establishing Theorem 1 and Theorem 4.

Lemma S3. *Suppose that assumptions (X2), (X4), (B1) and (A1)–(A8) are satisfied. Consider either a sparse design setting when* $n_i \leq N_0 < \infty$ *or a dense design when* $n_i = m \to \infty$, $i = 1, \ldots, n$. *Set* $a_n = a_{n1}$ *and* $b_n = b_{n1}$ *for the sparse case, and* $a_n = a_{n2}$ *and* $b_n = b_{n2}$ *for the dense case. For a new independent subject* i^*, *suppose that* $m^* = m^*(n) \to \infty$ *is such that* $m^*(a_n + b_n) = o(1)$ *as* $n \to \infty$. *If* $K = K(n)$ *satisfies* $(a_n + b_n) \sum_{k=1}^{K} \lambda_k^{-1} = o(1)$ *as* $n \to \infty$, *then*

$$\|\hat{\boldsymbol{\xi}}_K^* - \tilde{\boldsymbol{\xi}}_K^*\|_2^2 = O_p(R_n^*),$$

where

$$R_n^* = m^*(a_n + b_n)^2 \sum_{k=1}^{K} \delta_k^{-2} \lambda_k^{-2} + m^{*-1} \sum_{k=1}^{K} \lambda_k^{-2} + m^{*2}(a_n + b_n)^2 \sum_{k=1}^{K} \lambda_k^{-2} + m^{*4}(a_n + b_n)^4 \sum_{k=1}^{K} \delta_k^{-2} \lambda_k^{-2}.$$

Proof of Lemma S3. Similarly as in the proof of Theorem 1, write

$$\hat{\mathbf{e}}_k^* = \int_{\mathcal{T}} \hat{\Gamma}(\mathbf{T}^*, t) \hat{\phi}_k(t) dt - \hat{\boldsymbol{\Sigma}}^* \mathbf{W}^* \hat{\boldsymbol{\phi}}_k^*. \quad \text{(S.75)}$$

From Theorem 5.2 in Zhang and Wang (2016), we have

$$\|\hat{\Gamma} - \Gamma\|_\infty = O(a_n + b_n) \quad \text{a.s.}, \quad \text{(S.76)}$$

as $n \to \infty$, which implies

$$\|\hat{\Xi} - \Xi\|_{\text{op}} = O(a_n + b_n) \quad \text{a.s.}, \quad \text{(S.77)}$$

as $n \to \infty$. This combined with perturbation results (Bosq, 2000) show that for any $k \geq 1$,

$$\|\hat{\phi}_k - \phi_k\|_{L^2} \leq 2\sqrt{2} \delta_k^{-1} \|\hat{\Xi} - \Xi\|_{\text{op}} = O((a_n + b_n) \delta_k^{-1}) \quad \text{a.s.}, \quad \text{(S.78)}$$

and

$$|\hat{\lambda}_k - \lambda_k| \leq \|\hat{\Xi} - \Xi\|_{\text{op}} = O(a_n + b_n) \quad \text{a.s.}, \tag{S.79}$$

as $n \to \infty$. Similar to the proof of Theorem 2 in Dai *et al.* (2018) and employing Theorem 5.1 and 5.2 in Zhang and Wang (2016), it holds that

$$\|\hat{\mathbf{\Sigma}}^{*-1} - \mathbf{\Sigma}^{*-1}\|_{\text{op},2} \lesssim m^*(|\hat{\sigma}^2 - \sigma^2| + \|\hat{\Gamma} - \Gamma\|_\infty) = O(m^*(a_n + b_n)) \quad \text{a.s.}, \tag{S.80}$$

as $n \to \infty$. Also note that for $1 \leq k \leq K$,

$$\|\mathbf{W}^* \boldsymbol{\phi}_k^*\|_2 = O\left(\lambda_k^{-1} \left(\sum_{r=1}^{m^*} w_r^{*2}\right)^{1/2}\right). \tag{S.81}$$

Similar arguments as in the proof of Theorem 2 in Yao *et al.* (2005) along with perturbation results (Bosq, 2000), (S.76), and (S.78) show that

$$\begin{aligned}\sup_{t \in \mathcal{T}} |\hat{\lambda}_k \hat{\phi}_k(t) - \lambda_k \phi_k(t)| &\leq \|\hat{\Gamma} - \Gamma\|_\infty + \|\Gamma\|_\infty \|\hat{\phi}_k - \phi_k\|_{L^2} \\ &= O\left((a_n + b_n)(1 + \delta_k^{-1})\right) \quad \text{a.s.},\end{aligned} \tag{S.82}$$

as $n \to \infty$. By the Cauchy–Schwarz inequality and employing the orthonormality of the $\hat{\phi}_k$,

$$|\hat{\lambda}_k \hat{\phi}_k(t)| = \left| \int_{\mathcal{T}} \hat{\Gamma}(t,s) \hat{\phi}_k(s) ds \right| \leq \left(\int_{\mathcal{T}} \hat{\Gamma}^2(t,s) ds \right)^{1/2} \leq \|\hat{\Gamma}\|_\infty. \tag{S.83}$$

Since for large enough n we have

$$\lambda_K^{-1} \|\hat{\Xi} - \Xi\|_{\text{op}} \leq \sum_{k=1}^{K} \lambda_k^{-1} \|\hat{\Xi} - \Xi\|_{\text{op}} = O\left((a_n + b_n) \sum_{k=1}^{K} \lambda_k^{-1}\right) = o(1) \quad \text{a.s.},$$

where the first equality is due to (S.77) and the last is due to the condition $(a_n + b_n)\nu_K = o(1)$ as $n \to \infty$, we have $\|\hat{\Xi} - \Xi\|_{\text{op}} \le \lambda_K/2 \le \lambda_k/2$ a.s. for large enough n. In view of (S.79), it follows that for any $1 \le k \le K$,

$$|\hat{\lambda}_k - \lambda_k| \le \lambda_k/2 \quad \text{a.s.,} \tag{S.84}$$

as $n \to \infty$. Combining with (S.83) and (S.76) leads to

$$\|\hat{\phi}_k\|_\infty \le \hat{\lambda}_k^{-1}\|\hat{\Gamma}\|_\infty \le 2\lambda_k^{-1}(\|\hat{\Gamma} - \Gamma\|_\infty + \|\Gamma\|_\infty) = O(\lambda_k^{-1}) \quad \text{a.s.,} \tag{S.85}$$

for large enough n. This along with (S.79), (S.82), and (S.84) implies

$$\begin{aligned} \sup_{t\in\mathcal{T}} |\hat{\phi}_k(t) - \phi_k(t)| &\le \|\hat{\phi}_k\|_\infty \lambda_k^{-1}|\hat{\lambda}_k - \lambda_k| + \lambda_k^{-1}\|\hat{\lambda}_k\hat{\phi}_k - \lambda_k\phi_k\|_\infty \\ &= O\left((a_n + b_n)(\lambda_k^{-2} + \lambda_k^{-1} + \lambda_k^{-1}\delta_k^{-1})\right) \quad \text{a.s.,} \end{aligned} \tag{S.86}$$

as $n \to \infty$. Thus, using that $\delta_k \le \lambda_k$ we obtain

$$\|\mathbf{W}^*(\hat{\boldsymbol{\phi}}_k^* - \boldsymbol{\phi}_k^*)\|_2 = O\left(\left(\sum_{r=1}^{m^*} w_r^{*2}\right)^{1/2} \lambda_k^{-1}(a_n + b_n)(1 + \delta_k^{-1})\right) \quad \text{a.s.,} \tag{S.87}$$

as $n \to \infty$. Let $\boldsymbol{\phi}_k^* = \phi_k(\mathbf{T}^*)$ and $\hat{\boldsymbol{\phi}}_k^* = \hat{\phi}_k(\mathbf{T}^*)$. From (S.75), note that

$$\|\hat{\mathbf{e}}_k^*\|_2 \le \left\| \int_{\mathcal{T}} \hat{\Gamma}(\mathbf{T}^*, s)\hat{\phi}_k(s)ds - \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T})\mathbf{W}^*\hat{\boldsymbol{\phi}}_k^* \right\|_2 + \|\hat{\sigma}^2\mathbf{W}^*\hat{\boldsymbol{\phi}}_k^*\|_2,$$

where

$$\|\hat{\sigma}^2\mathbf{W}^*\hat{\boldsymbol{\phi}}_k^*\|_2^2 \le (|\hat{\sigma}^2 - \sigma^2| + \sigma^2)^2\|\hat{\phi}_k\|_\infty^2 \sum_{l=1}^{m^*} w_l^{*2} \lesssim \lambda_k^{-2}\sum_{l=1}^{m^*} w_l^{*2} \quad \text{a.s.,} \tag{S.88}$$

for large enough n and the last upper bound depends on k only through λ_k^{-2}. Here the last inequality uses that $\|\hat{\phi}_k\|_\infty = O(\lambda_k^{-1})$ a.s. and $|\hat{\sigma}^2 - \sigma^2| = O(a_n + b_n)$ a.s. as $n \to \infty$. Observe

$$\begin{aligned}\int_{\mathcal{T}} \hat{\Gamma}(\mathbf{T}^*, s)\hat{\phi}_k(s)ds - \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T})\mathbf{W}^*\hat{\boldsymbol{\phi}}_k^* &= \int_{\mathcal{T}} \hat{\Gamma}(\mathbf{T}^*, s)\hat{\phi}_k(s)ds - \int_{\mathcal{T}} \Gamma(\mathbf{T}^*, s)\phi_k(s)ds \\ &+ \int_{\mathcal{T}} \Gamma(\mathbf{T}^*, s)\phi_k(s)ds - \Gamma(\mathbf{T}^*, \mathbf{T}^{*T})\mathbf{W}^*\boldsymbol{\phi}_k^* \\ &+ \Gamma(\mathbf{T}^*, \mathbf{T}^{*T})\mathbf{W}^*\boldsymbol{\phi}_k^* - \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T})\mathbf{W}^*\hat{\boldsymbol{\phi}}_k^*,\end{aligned} \tag{S.89}$$

Hence, it suffices to control each of the differences in (S.89). First,

$$\begin{aligned}&\int_{\mathcal{T}} \hat{\Gamma}(\mathbf{T}^*, s)\hat{\phi}_k(s) - \Gamma(\mathbf{T}^*, s)\phi_k(s)ds \\ &= \int_{\mathcal{T}} (\hat{\Gamma}(\mathbf{T}^*, s) - \Gamma(\mathbf{T}^*, s))\hat{\phi}_k(s)ds + \int_{\mathcal{T}} \Gamma(\mathbf{T}^*, s)(\hat{\phi}_k(s) - \phi_k(s))ds,\end{aligned}$$

where, for $j = 1, \ldots, m^*$, and by using the orthonormality of the $\hat{\phi}_k$,

$$\begin{aligned}\left|\int_{\mathcal{T}} (\hat{\Gamma}(T_j^*, s) - \Gamma(T_j^*, s))\hat{\phi}_k(s)ds\right| &\leq \left(\int_{\mathcal{T}} (\hat{\Gamma}(T_j^*, s) - \Gamma(T_j^*, s))^2 ds\right)^{1/2} \\ &\leq \|\hat{\Gamma} - \Gamma\|_\infty \\ &= O(a_n + b_n) \quad \text{a.s.,}\end{aligned}$$

and

$$\left|\int_{\mathcal{T}} \Gamma(T_j^*, s)(\hat{\phi}_k(s) - \phi_k(s))ds\right| \leq \|\Gamma\|_\infty \|\hat{\phi}_k - \phi_k\|_{L^2} = O\left((a_n + b_n)\delta_k^{-1}\right) \quad \text{a.s.,}$$

where we use that $|\mathcal{T}| = 1$ and $\Gamma(s, t)$ is continuous over the compact set $\mathcal{T}^2$. Thus

$$\left\|\int_{\mathcal{T}} \hat{\Gamma}(\mathbf{T}^*, s)\hat{\phi}_k(s) - \Gamma(\mathbf{T}^*, s)\phi_k(s)ds\right\|_2 = O\left(\sqrt{m^*}(a_n + b_n)(1 + \delta_k^{-1})\right) \quad \text{a.s.,} \tag{S.90}$$

as $n \to \infty$, and the bound depends on k only through δ_k^{-1}. Second, from the Riemann sum approximation in (S.4) and noting that the application $g_j(t) = \Gamma(T_j^*, t)\phi_k(t)$ satisfies $\|g_j\|_\infty = O(\lambda_k^{-1})$ and $\|g_j'\|_\infty = O(\lambda_k^{-1})$ by (X3), where the $O(\lambda_k^{-1})$ terms are uniform in j and depend on k only through λ_k^{-1}, we have

$$\begin{aligned}&\left|\int_{\mathcal{T}} \Gamma(T_j^*, s)\phi_k(s)ds - \Gamma(T_j^*, \mathbf{T}^{*T})\mathbf{W}^*\boldsymbol{\phi}_k^*\right| \\ &\lesssim \lambda_k^{-1}\left(\sum_{l=1}^{m^*} w_l^{*2} + (1-\mathbf{T}^{(m^*)})^2 + (1-\mathbf{T}^{(m^*)})\right),\end{aligned}$$

where $\mathbf{T}^{(m^*)} := \max_{j=1,\dots,m^*} T_j^*$ and the upper bound is uniform in j and depends on k only through λ_k^{-1}. Thus

$$\begin{aligned}&\left\|\int_{\mathcal{T}} \Gamma(\mathbf{T}^*, s)\phi_k(s)ds - \Gamma(\mathbf{T}^*, \mathbf{T}^{*T})\mathbf{W}^*\boldsymbol{\phi}_k^*\right\|_2 \\ &= O\left(\sqrt{m^*}\lambda_k^{-1}\left(\sum_{l=1}^{m^*} w_l^{*2} + (1-\mathbf{T}^{(m^*)})^2 + (1-\mathbf{T}^{(m^*)})\right)\right). \qquad \text{(S.91)}\end{aligned}$$

Third, observe

$$\begin{aligned}\Gamma(\mathbf{T}^*, \mathbf{T}^{*T})\mathbf{W}^*\boldsymbol{\phi}_k^* - \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T})\mathbf{W}^*\hat{\boldsymbol{\phi}}_k^* &= (\Gamma(\mathbf{T}^*, \mathbf{T}^{*T}) - \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T}))\mathbf{W}^*\boldsymbol{\phi}_k^* \\ &\quad + \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T})(\mathbf{W}^*\boldsymbol{\phi}_k^* - \mathbf{W}^*\hat{\boldsymbol{\phi}}_k^*).\end{aligned} \qquad \text{(S.92)}$$

Note that

$$\begin{aligned}&\left\|(\Gamma(\mathbf{T}^*, \mathbf{T}^{*T}) - \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T}))\mathbf{W}^*\boldsymbol{\phi}_k^*\right\|_2 \\ &\le \left\|\Gamma(\mathbf{T}^*, \mathbf{T}^{*T}) - \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T})\right\|_{\text{op},2}\|\mathbf{W}^*\boldsymbol{\phi}_k^*\|_2 \\ &\lesssim \lambda_k^{-1}\left(\sum_{l=1}^{m^*} w_l^{*2}\right)^{1/2}\left\|\Gamma(\mathbf{T}^*, \mathbf{T}^{*T}) - \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T})\right\|_{\text{op},2},\end{aligned}$$

where the last equality follows similarly as in (S.7) and using that $\|\phi_k\|_\infty = O(\lambda_k^{-1})$. Since $\|A\|_{\text{op},2} \leq \|A\|_F$, where $\|A\|_F$ denotes the Frobenius norm of a squared matrix A, and

$$\left\|\Gamma(\mathbf{T}^*, \mathbf{T}^{*T}) - \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T})\right\|_F^2 \leq m^{*2} \sup_{s,t\in\mathcal{T}} |\hat{\Gamma}(s,t) - \Gamma(s,t)|^2 = O(m^{*2}(a_n + b_n)^2) \quad \text{a.s.,}$$

as $n \to \infty$, it follows that

$$\left\|(\Gamma(\mathbf{T}^*, \mathbf{T}^{*T}) - \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T}))\mathbf{W}^*\boldsymbol{\phi}_k^*\right\|_2 \lesssim \lambda_k^{-1} m^*(a_n + b_n) \left(\sum_{l=1}^{m^*} w_l^{*2}\right)^{1/2} \quad \text{a.s.,} \tag{S.93}$$

as $n \to \infty$. Also,

$$\begin{aligned}
&\left\|\hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T})(\mathbf{W}^*\boldsymbol{\phi}_k^* - \mathbf{W}^*\hat{\boldsymbol{\phi}}_k^*)\right\|_2 \\
&= \left(\left\|\hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T}) - \Gamma(\mathbf{T}^*, \mathbf{T}^{*T})\right\|_{\text{op},2} + \left\|\Gamma(\mathbf{T}^*, \mathbf{T}^{*T})\right\|_{\text{op},2}\right) \left\|\mathbf{W}^*(\boldsymbol{\phi}_k^* - \hat{\boldsymbol{\phi}}_k^*)\right\|_2 \\
&\lesssim (m^*(a_n + b_n) + m^*)\lambda_k^{-1}(a_n + b_n)(1 + \delta_k^{-1}) \left(\sum_{l=1}^{m^*} w_l^{*2}\right)^{1/2} \quad \text{a.s.} \\
&\lesssim m^*\lambda_k^{-1}(a_n + b_n)(1 + \delta_k^{-1}) \left(\sum_{l=1}^{m^*} w_l^{*2}\right)^{1/2} \quad \text{a.s.,}
\end{aligned}$$

as $n \to \infty$, where the first inequality follows from (S.87) and the last inequality uses the condition $m^*(a_n + b_n) = o(1)$ as $n \to \infty$. This along with (S.92) and (S.93) implies

$$\left\|\Gamma(\mathbf{T}^*, \mathbf{T}^{*T})\mathbf{W}^*\boldsymbol{\phi}_k^* - \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T})\mathbf{W}^*\hat{\boldsymbol{\phi}}_k^*\right\|_2 \lesssim m^*\lambda_k^{-1}(a_n + b_n)(1 + \delta_k^{-1}) \left(\sum_{l=1}^{m^*} w_l^{*2}\right)^{1/2}, \tag{S.94}$$

almost surely as $n \to \infty$, where the bound depends on k only through λ_k^{-1} and δ_k^{-1}. Combining (S.89), (S.90), (S.91), and (S.94) leads to

$$\begin{aligned}
&\left\| \int_{\mathcal{T}} \hat{\Gamma}(\mathbf{T}^*, t)\hat{\phi}_k(t)dt - \hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^{*T})\mathbf{W}^*\hat{\boldsymbol{\phi}}_k^* \right\|_2 \\
&\lesssim \sqrt{m^*}(a_n + b_n)(1 + \delta_k^{-1}) \\
&+ \sqrt{m^*}\lambda_k^{-1} \left(\sum_{l=1}^{m^*} w_l^{*2} + (1 - \mathbf{T}^{(m^*)})^2 + (1 - \mathbf{T}^{(m^*)}) \right) \\
&+ m^*\lambda_k^{-1}(a_n + b_n)(1 + \delta_k^{-1}) \left(\sum_{l=1}^{m^*} w_l^{*2} \right)^{1/2} \quad \text{a.s.,}
\end{aligned} \tag{S.95}$$

as $n \to \infty$. This along with (S.88) implies

$$\begin{aligned}
\|\hat{\mathbf{e}}_k^*\|_2 \lesssim{}& \sqrt{m^*}(a_n + b_n)(1 + \delta_k^{-1}) + \sqrt{m^*}\lambda_k^{-1} \left(\sum_{l=1}^{m^*} w_l^{*2} + (1 - \mathbf{T}^{(m^*)})^2 + (1 - \mathbf{T}^{(m^*)}) \right) \\
&+ m^*\lambda_k^{-1}(a_n + b_n)(1 + \delta_k^{-1}) \left(\sum_{l=1}^{m^*} w_l^{*2} \right)^{1/2} + \lambda_k^{-1} \left(\sum_{l=1}^{m^*} w_l^{*2} \right)^{1/2} \quad \text{a.s.,}
\end{aligned} \tag{S.96}$$

as $n \to \infty$, where the bound depends on k only through λ_k^{-1} and δ_k^{-1}. Define auxiliary quantities $Z_{m^*,n,K} := \sum_{k=1}^{K}[\hat{\mathbf{e}}_k^{*T}\hat{\boldsymbol{\Sigma}}^{*-1}(\mathbf{X}^* - \hat{\boldsymbol{\mu}}^*)]^2$, $\tilde{Z}_{m^*,n,K} := \sum_{k=1}^{K}[\hat{\mathbf{e}}_k^{*T}\hat{\boldsymbol{\Sigma}}^{*-1}(\mathbf{X}^* - \boldsymbol{\mu}^*)]^2$, $\mu_{m^*,n,K} := \sum_{k=1}^{K}[\hat{\mathbf{e}}_k^{*T}\hat{\boldsymbol{\Sigma}}^{*-1}(\boldsymbol{\mu}^* - \hat{\boldsymbol{\mu}}^*)]^2$, and observe

$$Z_{m^*,n,K} \lesssim \tilde{Z}_{m^*,n,K} + \mu_{m^*,n,K}. \tag{S.97}$$

By independence of the new subject's observations from the estimated population quantities, we have

$$E[Z_{m^*,n,K}|\mathbf{T}^*,\hat{\Gamma},\hat{\phi}_k,\hat{\sigma},\hat{\mu}] \lesssim E\Big[\tilde{Z}_{m^*,n,K}|\mathbf{T}^*,\hat{\Gamma},\hat{\phi}_k,\hat{\sigma},\hat{\mu}\Big] + \mu_{m^*,n,K}$$
$$= \sum_{k=1}^{K} \hat{\mathbf{e}}_k^{*T}\hat{\mathbf{\Sigma}}^{*-1}\mathbf{\Sigma}^*\hat{\mathbf{\Sigma}}^{*-1}\hat{\mathbf{e}}_k^* + \mu_{m^*,n,K} \quad \text{a.s.}, \quad \text{(S.98)}$$

and for large enough n

$$\Big|\sum_{k=1}^{K} \hat{\mathbf{e}}_k^{*T}\hat{\mathbf{\Sigma}}^{*-1}\mathbf{\Sigma}^*\hat{\mathbf{\Sigma}}^{*-1}\hat{\mathbf{e}}_k^*\Big|$$
$$\leq \Big|\sum_{k=1}^{K}[\hat{\mathbf{e}}_k^{*T}(\hat{\mathbf{\Sigma}}^{*-1}-\mathbf{\Sigma}^{*-1})\mathbf{\Sigma}^*(\hat{\mathbf{\Sigma}}^{*-1}-\mathbf{\Sigma}^{*-1})\hat{\mathbf{e}}_k^* + 2\hat{\mathbf{e}}_k^{*T}(\hat{\mathbf{\Sigma}}^{*-1}-\mathbf{\Sigma}^{*-1})\hat{\mathbf{e}}_k^* + \hat{\mathbf{e}}_k^{*T}\mathbf{\Sigma}^{*-1}\hat{\mathbf{e}}_k^*]\Big|$$
$$\lesssim \sum_{k=1}^{K}[m^{*3}(a_n+b_n)^2\|\hat{\mathbf{e}}_k^*\|_2^2 + m^*(a_n+b_n)\|\hat{\mathbf{e}}_k^*\|_2^2 + \|\hat{\mathbf{e}}_k^*\|_2^2] \quad \text{a.s.}$$
$$\lesssim (1+m^{*3}(a_n+b_n)^2)\sum_{k=1}^{K}\|\hat{\mathbf{e}}_k^*\|_2^2 \quad \text{a.s.}$$
$$\lesssim [m^*(a_n+b_n)^2\left(\sum_{k=1}^{K}\delta_k^{-2}\right) + m^*\left(\sum_{k=1}^{K}\lambda_k^{-2}\right)\left(\sum_{l=1}^{m^*}w_l^{*2} + (1-\mathbf{T}^{(m^*)})^2 + (1-\mathbf{T}^{(m^*)})\right)^2$$
$$+ m^{*2}(a_n+b_n)^2\left(\sum_{k=1}^{K}\lambda_k^{-2}\delta_k^{-2}\right)\sum_{l=1}^{m^*}w_l^{*2} + \left(\sum_{k=1}^{K}\lambda_k^{-2}\right)\sum_{l=1}^{m^*}w_l^{*2}](1+m^{*3}(a_n+b_n)^2)$$
$$= (1+m^{*3}(a_n+b_n)^2)$$
$$O_p\left(m^*(a_n+b_n)^2\left(\sum_{k=1}^{K}\delta_k^{-2}\right) + m^{*-1}\left(\sum_{k=1}^{K}\lambda_k^{-2}\right) + m^*(a_n+b_n)^2\left(\sum_{k=1}^{K}\lambda_k^{-2}\delta_k^{-2}\right)\right)$$
$$= (1+m^{*3}(a_n+b_n)^2)O_p\left(m^{*-1}\left(\sum_{k=1}^{K}\lambda_k^{-2}\right) + m^*(a_n+b_n)^2\left(\sum_{k=1}^{K}\lambda_k^{-2}\delta_k^{-2}\right)\right)$$
$$= O_p(R_n^*), \quad \text{(S.99)}$$

where the second inequality is due to $\|\hat{\mathbf{\Sigma}}^{*-1}-\mathbf{\Sigma}^{*-1}\|_{\text{op},2} = O(m^*(a_n+b_n))$ a.s. as $n \to \infty$, $\|\mathbf{\Sigma}^{*-1}\|_{\text{op},2} \leq \sigma^{-2}$, $\|\mathbf{\Sigma}^*\|_{\text{op},2} = O(m^*)$, and the fourth inequality follows from (S.96). This shows that

$$\sum_{k=1}^{K} \hat{\mathbf{e}}_k^{*T} \hat{\mathbf{\Sigma}}^{*-1} \mathbf{\Sigma}^* \hat{\mathbf{\Sigma}}^{*-1} \hat{\mathbf{e}}_k^* = O_p(R_n^*).$$

Thus, for any $\epsilon > 0$ there exists $N_0 = N_0(\epsilon) \geq 1$ and $M_0 = M_0(\epsilon) > 0$ such that for all $n \geq N_0$

$$P\left(R_n^{*-1} \left| \sum_{k=1}^{K} \hat{\mathbf{e}}_k^{*T} \hat{\mathbf{\Sigma}}^{*-1} \mathbf{\Sigma}^* \hat{\mathbf{\Sigma}}^{*-1} \hat{\mathbf{e}}_k^* \right| > M_0 \right) \leq \epsilon. \tag{S.100}$$

Let $M > 0$ and define

$$u_{m^*,n,K} = P\left(R_n^{*-1} \tilde{Z}_{m^*,n,K} > M | \mathbf{T}^*, \hat{\Gamma}, \hat{\phi}_k, \hat{\sigma}, \hat{\mu}\right).$$

Choosing $M = M(\epsilon) = M_0/\epsilon$ and using that $u_{m^*,n,K} \leq 1$ along with the relation

$$u_{m^*,n,K} \lesssim \frac{1}{R_n^* M} \sum_{k=1}^{K} \hat{\mathbf{e}}_k^{*T} \hat{\mathbf{\Sigma}}^{*-1} \mathbf{\Sigma}^* \hat{\mathbf{\Sigma}}^{*-1} \hat{\mathbf{e}}_k^*,$$

which follows analogously as in (S.98), leads to

$$\begin{aligned} P\left(R_n^{*-1} \tilde{Z}_{m^*,n,K} > M\right) &= E(u_{m^*,n,K} 1_{\{u_{m^*,n,K} \leq \epsilon\}} + u_{m^*,n,K} 1_{\{u_{m^*,n,K} > \epsilon\}}) \\ &\leq \epsilon + P(u_{m^*,n,K} > \epsilon) \\ &\leq 2\epsilon, \end{aligned}$$

where the last inequality follows from (S.100). Therefore

$$\tilde{Z}_{m^*,n,K} = O_p(R_n^*).$$

Also, for large enough n and using (S.96) along with $\|\hat{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^*\|_2^2 = O(m^*(a_n + b_n)^2)$ a.s., we obtain

$$\mu_{m^*,n,K} \lesssim m^*(a_n + b_n)^2 \sum_{k=1}^{K} \|\hat{\mathbf{e}}_k^*\|_2^2 \quad \text{a.s.},$$

which in view of the third inequality in (S.99) and the condition $m^*(a_n + b_n) = o(1)$ as $n \to \infty$ is of slower order compared to the rate $O_p(R_n^*)$. These along with (S.97) leads to

$$Z_{m^*,n,K} = O_p(R_n^*). \tag{S.101}$$

Then a conditioning argument leads to

$$\begin{aligned}
E[(\mathbf{e}_k^{*T}\boldsymbol{\Sigma}^{*-1}(\mathbf{X}^* - \boldsymbol{\mu}^*))^2] &= E(E[(\mathbf{e}_k^{*T}\boldsymbol{\Sigma}^{*-1}(\mathbf{X}^* - \boldsymbol{\mu}^*))^2|\mathbf{T}^*]) \\
&= E(\operatorname{Var}[\mathbf{e}_k^{*T}\boldsymbol{\Sigma}^{*-1}(\mathbf{X}^* - \boldsymbol{\mu}^*)|\mathbf{T}^*]) \\
&= E(\mathbf{e}_k^{*T}\boldsymbol{\Sigma}^{*-1}\mathbf{e}_k^*) \\
&\leq \sigma^{-2}E(\|\mathbf{e}_k^*\|_2^2) \\
&\lesssim m^{*-1}\lambda_k^{-2},
\end{aligned}$$

where the last inequality holds for large enough n and follows analogously as in (S.21). This implies

$$E\Big[\sum_{k=1}^{K}(\mathbf{e}_k^{*T}\boldsymbol{\Sigma}^{*-1}(\mathbf{X}^* - \boldsymbol{\mu}^*))^2\Big] \leq m^{*-1}\sum_{k=1}^{K}\lambda_k^{-2}.$$

Hence

$$\sum_{k=1}^{K}(\mathbf{e}_k^{*T}\boldsymbol{\Sigma}^{*-1}(\mathbf{X}^* - \boldsymbol{\mu}^*))^2 = O_p\left(m^{*-1}\sum_{k=1}^{K}\lambda_k^{-2}\right). \tag{S.102}$$

For any $k = 1, \ldots, K$, observe

$$\begin{aligned}
&\hat{\xi}_k^* - \tilde{\xi}_k^* \\
&= \hat{\mathbf{e}}_k^{*T}\hat{\boldsymbol{\Sigma}}^{*-1}(\mathbf{X}^* - \hat{\boldsymbol{\mu}}^*) + \hat{\boldsymbol{\phi}}_k^{*T}\mathbf{W}^*(\mathbf{X}^* - \hat{\boldsymbol{\mu}}^*) - \mathbf{e}_k^{*T}\boldsymbol{\Sigma}^{*-1}(\mathbf{X}^* - \boldsymbol{\mu}^*) - \boldsymbol{\phi}_k^{*T}\mathbf{W}^*(\mathbf{X}^* - \boldsymbol{\mu}^*).
\end{aligned} \tag{S.103}$$

From (S.81), (S.87), and using that $\|\mathbf{X}^* - \boldsymbol{\mu}^*\|_2^2 = O_p(m^*)$ and $\|\hat{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^*\|_2^2 = O_p(m^*(a_n + b_n)^2)$, we obtain

$$\begin{aligned}
&\sum_{k=1}^{K}[(\hat{\boldsymbol{\phi}}_k^{*T}\mathbf{W}^*(\mathbf{X}^* - \hat{\boldsymbol{\mu}}^*) - \boldsymbol{\phi}_k^{*T}\mathbf{W}^*(\mathbf{X}^* - \boldsymbol{\mu}^*))^2] \\
&\lesssim \sum_{k=1}^{K}[\|\mathbf{W}^*(\hat{\boldsymbol{\phi}}_k^* - \boldsymbol{\phi}_k^*)\|_2^2\|\mathbf{X}^* - \hat{\boldsymbol{\mu}}^*\|_2^2 + \|\mathbf{W}^*\boldsymbol{\phi}_k^*\|_2^2\|\hat{\boldsymbol{\mu}}^* - \boldsymbol{\mu}^*\|_2^2] \\
&= O_p\left((a_n + b_n)^2 \sum_{k=1}^{K} \lambda_k^{-2}\delta_k^{-2}\right). \qquad \text{(S.104)}
\end{aligned}$$

Combining (S.101), (S.102), (S.103), and (S.104) leads to

$$\|\hat{\boldsymbol{\xi}}_K^* - \tilde{\boldsymbol{\xi}}_K^*\|_2^2 = \sum_{k=1}^{K}[(\hat{\xi}_k^* - \tilde{\xi}_k^*)^2] = O_p(R_n^*),$$

which shows the result. □

Lemma S4. *Suppose that assumptions (X2), (X4), (B1) and (A1)–(A8) are satisfied. Consider either a sparse design setting when $n_i \le N_0 < \infty$ or a dense design when $n_i = m \to \infty$, $i = 1, \ldots, n$. Set $a_n = a_{n1}$ and $b_n = b_{n1}$ for the sparse case, and $a_n = a_{n2}$ and $b_n = b_{n2}$ for the dense case. Let $\upsilon_K = \sum_{k=1}^{K} \lambda_k^{-1/2}\delta_k^{-1}$. For a new independent subject i^*, suppose that $m^* = m^*(n) \to \infty$ is such that $m^*(a_n + b_n) = o(1)$ and $K = K(n)$ satisfies $(a_n + b_n)\upsilon_K = o(1)$ as $n \to \infty$. Then*

$$\mathit{trace}(\hat{\boldsymbol{\Sigma}}_K^* - \boldsymbol{\Sigma}_K^*) = O_p\left(m^*(a_n + b_n)^2 \sum_{k=1}^{K} \lambda_k^{-2}\delta_k^{-2} + (a_n + b_n)\sum_{k=1}^{K} \lambda_k^{-2}\delta_k^{-1}\right).$$

Proof of Lemma S4. In effect, for $j = 1, \ldots, K$, the (j, j)-element of $\hat{\boldsymbol{\Sigma}}_K^* - \boldsymbol{\Sigma}_K^*$

is given by

$$[\hat{\mathbf{\Sigma}}_K^* - \mathbf{\Sigma}_K^*]_{j,j} = \hat{\mathbf{e}}_j^{*T}\hat{\mathbf{\Sigma}}^{*-1}\hat{\mathbf{e}}_j^* + \hat{\mathbf{e}}_j^{*T}\mathbf{W}^*\hat{\boldsymbol{\phi}}_j^* + \hat{\boldsymbol{\phi}}_j^{*T}\mathbf{W}^*\hat{\mathbf{e}}_j^* + \hat{\boldsymbol{\phi}}_j^{*T}\mathbf{W}^*\hat{\mathbf{\Sigma}}^*\mathbf{W}^*\hat{\boldsymbol{\phi}}_j^*$$
$$- (\mathbf{e}_j^{*T}\mathbf{\Sigma}^{*-1}\mathbf{e}_j^* + \mathbf{e}_j^{*T}\mathbf{W}^*\boldsymbol{\phi}_j^* + \boldsymbol{\phi}_j^{*T}\mathbf{W}^*\mathbf{e}_j^* + \boldsymbol{\phi}_j^{*T}\mathbf{W}^*\mathbf{\Sigma}^*\mathbf{W}^*\boldsymbol{\phi}_j^*), \tag{S.105}$$

where $\hat{\mathbf{e}}_j^*$ is defined as in (S.75). Note that the conditions of Lemma S3 hold since $(a_n + b_n)\nu_K = o(1)$ which is due to $\nu_K \leq \upsilon_K$ and $\delta_k \leq \lambda_k$, where $\nu_K = \sum_{k=1}^K \lambda_k^{-1}$. Observing for any $k = 1, \dots, K$,

$$\delta_k^{-1} \leq \sum_{k=1}^K \delta_k^{-1} = \sum_{k=1}^K \lambda_k^{-1/2}\delta_k^{-1}\lambda_k^{1/2} \leq \lambda_1^{1/2}\sum_{k=1}^K \lambda_k^{-1/2}\delta_k^{-1},$$

along with the condition $\upsilon_K(a_n + b_n) = o(1)$ as $n \to \infty$ leads to

$$\delta_k^{-1}(a_n + b_n) \leq \lambda_1^{1/2}(a_n + b_n)\sum_{k=1}^K \lambda_k^{-1/2}\delta_k^{-1} = o(1),$$

as $n \to \infty$, where the bound is uniform in k. This along with (S.81) and (S.87) imply

$$\|\mathbf{W}^*\hat{\boldsymbol{\phi}}_k^*\|_2 \leq \|\mathbf{W}^*(\hat{\boldsymbol{\phi}}_k^* - \boldsymbol{\phi}_k^*)\|_2 + \|\mathbf{W}^*\boldsymbol{\phi}_k^*\|_2 = O\left(\lambda_k^{-1}\left(\sum_{r=1}^{m^*} w_r^{*2}\right)^{1/2}\right) \quad \text{a.s.}, \tag{S.106}$$

as $n \to \infty$, where the bound depends on k only through λ_k^{-1}. Also, using (S.76) and since $m^*(a_n + b_n) = o(1)$ and $|\hat{\sigma}^2 - \sigma^2| = O(a_n + b_n)$ as $n \to \infty$, which follows from Proposition 1 in Dai *et al.* (2018), we obtain

$$\|\hat{\mathbf{\Sigma}}^*\|_{\text{op},2} \leq \hat{\sigma}^2 + \|\hat{\Gamma}(\mathbf{T}^*, \mathbf{T}^*)\|_{\text{op},2} \leq \hat{\sigma}^2 + m^*\|\hat{\Gamma}\|_\infty = O(m^*) \quad \text{a.s.}, \tag{S.107}$$

as $n \to \infty$. This along with (S.87), (S.106), and (S.107) leads to

$$\begin{aligned} |(\hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\phi}_j^*)^T \mathbf{W}^* \hat{\boldsymbol{\Sigma}}^* \mathbf{W}^* \hat{\boldsymbol{\phi}}_j^*| &\leq \|\mathbf{W}^*(\hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\phi}_j^*)\|_2 \|\hat{\boldsymbol{\Sigma}}^* \mathbf{W}^* \hat{\boldsymbol{\phi}}_j^*\|_2 \\ &\leq \|\mathbf{W}^*(\hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\phi}_j^*)\|_2 \|\hat{\boldsymbol{\Sigma}}^*\|_{\text{op},2} \|\mathbf{W}^* \hat{\boldsymbol{\phi}}_j^*\|_2 \\ &= O\left(\left(\sum_{r=1}^{m^*} w_r^{*2}\right) m^*(a_n + b_n)\lambda_j^{-2}(1 + \delta_j^{-1})\right) \quad \text{a.s.,} \end{aligned} \tag{S.108}$$

as $n \to \infty$, where the bound depends on j only through λ_j^{-1} and δ_j^{-1}. Using the fact that $\|\hat{\boldsymbol{\Sigma}}^* - \boldsymbol{\Sigma}^*\|_{\text{op},2} = O(m(a_n + b_n))$ a.s. as $n \to \infty$ along with (S.87) and (S.106), we obtain

$$\begin{aligned} \|\hat{\boldsymbol{\Sigma}}^* \mathbf{W}^* \hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\Sigma}^* \mathbf{W}^* \boldsymbol{\phi}_j^*\|_2 &\leq \|\hat{\boldsymbol{\Sigma}}^* - \boldsymbol{\Sigma}^*\|_{\text{op},2} \|\mathbf{W}^* \hat{\boldsymbol{\phi}}_j^*\|_2 + \|\boldsymbol{\Sigma}^*\|_{\text{op},2} \|\mathbf{W}^*(\hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\phi}_j^*)\|_2 \\ &= O\left(m^*(a_n + b_n)\lambda_j^{-1} \left(\sum_{r=1}^{m^*} w_r^{*2}\right)^{1/2} (1 + \delta_j^{-1})\right) \quad \text{a.s.,} \end{aligned} \tag{S.109}$$

as $n \to \infty$, where the bound depends on j only through λ_j^{-1} and δ_j^{-1}. Thus

$$\begin{aligned} |\boldsymbol{\phi}_j^{*T} \mathbf{W}^* (\hat{\boldsymbol{\Sigma}}^* \mathbf{W}^* \hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\Sigma}^* \mathbf{W}^* \boldsymbol{\phi}_j^*)| &\leq \|\mathbf{W}^* \boldsymbol{\phi}_j^*\|_2 \|\hat{\boldsymbol{\Sigma}}^* \mathbf{W}^* \hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\Sigma}^* \mathbf{W}^* \boldsymbol{\phi}_j^*\|_2 \\ &= O\left(m^*(a_n + b_n)\lambda_j^{-2} \left(\sum_{r=1}^{m^*} w_r^{*2}\right) (1 + \delta_j^{-1})\right) \quad \text{a.s.,} \end{aligned} \tag{S.110}$$

as $n \to \infty$, where the bound depends on j only through λ_j^{-2} and δ_j^{-1}. This combined with (S.108) leads to

$$
\begin{aligned}
&|\hat{\boldsymbol{\phi}}_j^{*T}\mathbf{W}^*\hat{\boldsymbol{\Sigma}}^*\mathbf{W}^*\hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\phi}_j^{*T}\mathbf{W}^*\boldsymbol{\Sigma}^*\mathbf{W}^*\boldsymbol{\phi}_j^*| \\
&\leq |(\hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\phi}_j^*)^T\mathbf{W}^*\hat{\boldsymbol{\Sigma}}^*\mathbf{W}^*\hat{\boldsymbol{\phi}}_j^*| + |\boldsymbol{\phi}_j^{*T}\mathbf{W}^*(\hat{\boldsymbol{\Sigma}}^*\mathbf{W}^*\hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\Sigma}^*\mathbf{W}^*\boldsymbol{\phi}_j^*)| \\
&= O\left(m^*(a_n+b_n)\lambda_j^{-2}\left(\sum_{r=1}^{m^*} w_r^2\right)(1+\delta_j^{-1})\right) \quad \text{a.s.,} \qquad \text{(S.111)}
\end{aligned}
$$

as $n \to \infty$, where the bound depends on j only through λ_j^{-2} and δ_j^{-1}. Write $\hat{\Delta}_k = \hat{\mathbf{e}}_k^* - \mathbf{e}_k^*$, $k = 1, \ldots, K$, and observe

$$
\|\hat{\lambda}_j\hat{\boldsymbol{\phi}}_j^* - \lambda_j\boldsymbol{\phi}_j^*\|_2 \leq m^{*1/2}\|\hat{\lambda}_j\hat{\phi}_j - \lambda_j\phi_j\|_\infty = O\left(m^{*1/2}(a_n+b_n)(1+\delta_j^{-1})\right) \quad \text{a.s.,}
$$

as $n \to \infty$, where the last equality is due to (S.82) and the bound depends on j only through δ_j^{-1}. This along with (S.109) leads to

$$
\begin{aligned}
\|\hat{\Delta}_j\|_2 &= \left\| \int_{\mathcal{T}} \hat{\Gamma}(\mathbf{T}^*, s)\hat{\phi}_j(s) - \Gamma(\mathbf{T}^*, s)\phi_j(s)ds + \boldsymbol{\Sigma}^*\mathbf{W}^*\boldsymbol{\phi}_j^* - \hat{\boldsymbol{\Sigma}}^*\mathbf{W}^*\hat{\boldsymbol{\phi}}_j^* \right\|_2 \\
&\leq \|\hat{\lambda}_j\hat{\boldsymbol{\phi}}_j^* - \lambda_j\boldsymbol{\phi}_j^*\|_2 + \|\hat{\boldsymbol{\Sigma}}^*\mathbf{W}^*\hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\Sigma}^*\mathbf{W}^*\boldsymbol{\phi}_j^*\|_2 \\
&= O\left(m^{*1/2}(a_n+b_n)(1+\delta_j^{-1})\left(1 + m^{*1/2}\lambda_j^{-1}\left(\sum_{r=1}^{m^*} w_r^{*2}\right)^{1/2}\right)\right) \quad \text{a.s.,}
\end{aligned}
$$

(S.112)

as $n \to \infty$, where the bound depends on j only through λ_j^{-1} and δ_j^{-1}. Using that $m^*(a_n+b_n) = o(1)$ along with $\|\hat{\boldsymbol{\Sigma}}^{*-1} - \boldsymbol{\Sigma}^{*-1}\|_{\text{op},2} = O(m^*(a_n+b_n))$ a.s. as $n \to \infty$, observe

$$
|\hat{\mathbf{e}}_j^{*T}\hat{\boldsymbol{\Sigma}}^{*-1}\hat{\mathbf{e}}_j^* - \mathbf{e}_j^{*T}\boldsymbol{\Sigma}^{*-1}\mathbf{e}_j^*| = O\left(\|\hat{\Delta}_j\|_2^2 + \|\hat{\Delta}_j\|_2\|\mathbf{e}_j^*\|_2 + m^*(a_n+b_n)\|\mathbf{e}_j^*\|_2^2\right) \quad \text{a.s.,}
$$

(S.113)

as $n \to \infty$, where the bound depends on j only through $\|\hat{\Delta}_j\|_2$ and $\|\mathbf{e}_j^*\|_2$. Also,

$$|\hat{\mathbf{e}}_j^{*T}\mathbf{W}^*\hat{\boldsymbol{\phi}}_j^* - \mathbf{e}_j^{*T}\mathbf{W}^*\boldsymbol{\phi}_j^*| \leq \|\hat{\Delta}_j\|_2\|\mathbf{W}^*\hat{\boldsymbol{\phi}}_j^*\|_2 + \|\mathbf{e}_j^*\|_2\|\mathbf{W}^*(\hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\phi}_j^*)\|_2. \tag{S.114}$$

For large enough n and in view of (S.111), (S.112), and using that the bound (S.20) holds analogously for $\|\mathbf{e}_j^*\|_2$ and the time points $\mathbf{T}^*$, we obtain

$$\sum_{j=1}^{K}[\hat{\boldsymbol{\phi}}_j^{*T}\mathbf{W}^*\hat{\boldsymbol{\Sigma}}^*\mathbf{W}^*\hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\phi}_j^{*T}\mathbf{W}^*\boldsymbol{\Sigma}^*\mathbf{W}^*\boldsymbol{\phi}_j^*] = O_p\left((a_n + b_n)\sum_{k=1}^{K}\lambda_k^{-2}\delta_k^{-1}\right), \tag{S.115}$$

and

$$\sum_{j=1}^{K}[\|\hat{\Delta}_j\|_2\|\mathbf{e}_j^*\|_2] = O_p\left((a_n + b_n)\sum_{j=1}^{K}\lambda_j^{-2}\delta_j^{-1}\right), \tag{S.116}$$

and

$$\sum_{j=1}^{K}[\|\hat{\Delta}_j\|_2^2] = O_p\left(m^*(a_n + b_n)^2\sum_{j=1}^{K}\lambda_j^{-2}\delta_j^{-2}\right). \tag{S.117}$$

Since $E(\|\mathbf{e}_j^*\|_2^2) \lesssim m^{*-1}\lambda_j^{-2}$, which follows analogously as in (S.21), we also have

$$\sum_{j=1}^{K}\|\mathbf{e}_j^*\|_2^2 = O_p\left(m^{*-1}\sum_{j=1}^{K}\lambda_j^{-2}\right). \tag{S.118}$$

From (S.106) and (S.112), we obtain

$$\sum_{j=1}^{K}[\|\hat{\Delta}_j\|_2\|\mathbf{W}^*\hat{\boldsymbol{\phi}}_j^*\|_2] = O_p\left((a_n + b_n)\sum_{j=1}^{K}\lambda_j^{-2}\delta_j^{-1}\right), \tag{S.119}$$

and using (S.87) we also have

$$\sum_{j=1}^{K}[\|\mathbf{e}_j^*\|_2\|\mathbf{W}^*(\hat{\boldsymbol{\phi}}_j^* - \boldsymbol{\phi}_j^*)\|_2] = O_p\left((a_n + b_n)m^{*-1}\sum_{j=1}^{K}\lambda_j^{-2}\delta_j^{-1}\right). \tag{S.120}$$

Combining (S.113), (S.116), (S.117), and (S.118) implies

$$\begin{aligned}
&\sum_{j=1}^{K}|\hat{\mathbf{e}}_j^{*T}\hat{\boldsymbol{\Sigma}}^{*-1}\hat{\mathbf{e}}_j^{*}-\mathbf{e}_j^{*T}\boldsymbol{\Sigma}^{*-1}\mathbf{e}_j^{*}| \\
&\lesssim \sum_{j=1}^{K}[\|\hat{\Delta}_j\|_2^2+\|\hat{\Delta}_j\|_2\|\mathbf{e}_j^{*}\|_2+m^{*}(a_n+b_n)\|\mathbf{e}_j^{*}\|_2^2] \\
&= O_p\left(m^{*}(a_n+b_n)^2\sum_{j=1}^{K}\lambda_j^{-2}\delta_j^{-2}+(a_n+b_n)\sum_{j=1}^{K}\lambda_j^{-2}\delta_j^{-1}\right), \qquad \text{(S.121)}
\end{aligned}$$

while combining (S.114), (S.119), and (S.120) leads to

$$\sum_{j=1}^{K}|\hat{\mathbf{e}}_j^{*T}\mathbf{W}^{*}\hat{\boldsymbol{\phi}}_j^{*}-\mathbf{e}_j^{*T}\mathbf{W}^{*}\boldsymbol{\phi}_j^{*}| = O_p\left((a_n+b_n)\sum_{j=1}^{K}\lambda_j^{-2}\delta_j^{-1}\right). \qquad \text{(S.122)}$$

Combining (S.105), (S.115), (S.121), and (S.122) leads to

$$\begin{aligned}
&|\text{trace}(\hat{\boldsymbol{\Sigma}}_K^{*}-\boldsymbol{\Sigma}_K^{*})| \\
&\leq \sum_{j=1}^{K}|[\hat{\boldsymbol{\Sigma}}_K^{*}-\boldsymbol{\Sigma}_K^{*}]_{j,j}| \ = O_p\left(m^{*}(a_n+b_n)^2\sum_{j=1}^{K}\lambda_j^{-2}\delta_j^{-2}+(a_n+b_n)\sum_{j=1}^{K}\lambda_j^{-2}\delta_j^{-1}\right),
\end{aligned}$$

and the result follows. □

For the following, recall that $w_i := \left(\sum_{l=1}^{n} n_l\right)^{-1}$, $\upsilon_M = \sum_{m=1}^{M}\delta_m^{-1}$ and $C(t) = E((X(t)-\mu(t))Y) = \int_{\mathcal{T}}\beta(s)\Gamma(t,s)ds, t \in \mathcal{T}$.

Lemma S5. *Suppose that (X4), (B1)-(B4), (A1)–(A8) hold and consider a sparse design with $n_i \leq N_0 < \infty$, setting $a_n = a_{n1}$ and $b_n = b_{n1}$. Then*

$$n^{-1}\sum_{i=1}^{n}\|\hat{\tilde{\boldsymbol{\xi}}}_{iK}-\tilde{\boldsymbol{\xi}}_{iK}\|_2^2 = O_p((a_n+b_n)^2), \qquad \text{(S.123)}$$

and

$$n^{-1}\sum_{i=1}^{n}\|\tilde{\boldsymbol{\xi}}_{iK}\|_2^2 = O_p(1). \qquad \text{(S.124)}$$

Proof of Lemma S5. First note that $\|\hat{\mu}-\mu\|_\infty = O(a_n)$ a.s. and $\|\hat{\Gamma}-\Gamma\|_\infty = O(a_n+b_n)$ a.s., which are due to Theorem 5.1 and 5.2 in Zhang and Wang (2016). From arguments in the proof of Theorem 2 in Dai *et al.* (2018) and noting that the constant c that appears in Lemma A.3 in Facer and Müller (2003) can be taken as a universal constant $c=2$,

$$\|\hat{\boldsymbol{\xi}}_{iK}-\tilde{\boldsymbol{\xi}}_{iK}\|_2^2 \le O((a_n+b_n)^2)\|\mathbf{X}_i-\hat{\boldsymbol{\mu}}_i\|_2^2 + O(a_n^2) + O(a_n(a_n+b_n))\|\mathbf{X}_i-\hat{\boldsymbol{\mu}}_i\|_2 \quad \text{a.s.,} \tag{S.125}$$

where the $O((a_n+b_n)^2)$, $O(a_n^2)$ and $O(a_n(a_n+b_n))$ terms are uniform in i. Let $\mathbf{U}_i = (X_i(T_{i1}),\ldots,X_i(T_{in_i}))^T$ be the true but unobserved values of the trajectory for the ith subject at the time points $\mathbf{T}_i$, so that by construction $\mathbf{X}_i = \mathbf{U}_i+\boldsymbol{\epsilon}_i$. Then

$$\begin{aligned} n^{-1}\sum_{i=1}^n \|\mathbf{X}_i-\hat{\boldsymbol{\mu}}_i\|_2 &= n^{-1}\sum_{i=1}^n \|\mathbf{U}_i+\boldsymbol{\epsilon}_i-\hat{\boldsymbol{\mu}}_i\|_2 \\ &\le n^{-1}\sum_{i=1}^n \|\mathbf{U}_i-\boldsymbol{\mu}_i\|_2 + n^{-1}\sum_{i=1}^n \|\boldsymbol{\epsilon}_i\|_2 + n^{-1}\sum_{i=1}^n \|\boldsymbol{\mu}_i-\hat{\boldsymbol{\mu}}_i\|_2, \end{aligned} \tag{S.126}$$

where $n^{-1}\sum_{i=1}^n \|\boldsymbol{\mu}_i-\hat{\boldsymbol{\mu}}_i\|_2 = O(a_n)$ almost surely. Since $n_i \le N_0$ in the sparse case, it is easy to show that $n^{-1}\sum_{i=1}^n \|\boldsymbol{\epsilon}_i\|_2 = O_p(1)$ and by Jensen's inequality

$$\begin{aligned} E\left(n^{-1}\sum_{i=1}^n \|\mathbf{U}_i-\boldsymbol{\mu}_i\|_2\right) &\le n^{-1}\sum_{i=1}^n \left(\sum_{j=1}^{n_i} E(X_i(T_{ij})-\mu(T_{ij}))^2\right)^{1/2} \\ &= n^{-1}\sum_{i=1}^n \left(\sum_{j=1}^{n_i} E(\Gamma(T_{ij},T_{ij}))\right)^{1/2} \le (\|\Gamma\|_\infty N_0)^{1/2} = O(1), \end{aligned}$$

where the first equality follow by conditioning on T_{ij}. This shows that $n^{-1}\sum_{i=1}^n \|\mathbf{U}_i-$

$\boldsymbol{\mu}_i\|_2 = O_p(1)$. Combining with (S.126) leads to

$$n^{-1}\sum_{i=1}^{n}\|\mathbf{X}_i - \hat{\boldsymbol{\mu}}_i\|_2 = O_p(1). \qquad \text{(S.127)}$$

By the triangle inequality

$$\|\mathbf{X}_i - \hat{\boldsymbol{\mu}}_i\|_2^2 \le \|\mathbf{U}_i - \boldsymbol{\mu}_i\|_2^2 + \|\boldsymbol{\epsilon}_i\|_2^2 + \|\boldsymbol{\mu}_i - \hat{\boldsymbol{\mu}}_i\|_2^2$$
$$+ 2\|\mathbf{U}_i - \boldsymbol{\mu}_i\|_2\|\boldsymbol{\epsilon}_i\|_2 + 2\|\mathbf{U}_i - \boldsymbol{\mu}_i\|_2\|\boldsymbol{\mu}_i - \hat{\boldsymbol{\mu}}_i\|_2 + 2\|\boldsymbol{\epsilon}_i\|_2\|\boldsymbol{\mu}_i - \hat{\boldsymbol{\mu}}_i\|_2,$$

where $\|\boldsymbol{\mu}_i - \hat{\boldsymbol{\mu}}_i\|_2 \le \sqrt{N_0 \sup_{t\in\mathcal{T}}(\mu(t) - \hat{\mu}(t))^2} = O(a_n)$ a.s. and uniformly over i. This along with the independence of $\boldsymbol{\epsilon}_i$ and $\mathbf{U}_i$, conditionally on $\mathbf{T}_i$, and using similar arguments as before, leads to $E\|\mathbf{X}_i - \hat{\boldsymbol{\mu}}_i\|_2^2 = O(1)$ uniformly over i. Thus

$$n^{-1}\sum_{i=1}^{n}\|\mathbf{X}_i - \hat{\boldsymbol{\mu}}_i\|_2^2 = O_p(1). \qquad \text{(S.128)}$$

Combining (S.125), (S.127) and (S.128) leads to the first result in (S.123). Note that

$$E(\tilde{\boldsymbol{\xi}}_{iK}^T\tilde{\boldsymbol{\xi}}_{iK})^2 \le E(\left\|\boldsymbol{\Lambda}_K\boldsymbol{\Phi}_{iK}^T\boldsymbol{\Sigma}_i^{-1}\right\|_{\text{op},2}^4 E(\|\mathbf{X}_i - \boldsymbol{\mu}_i\|_2^4\,|\mathbf{T}_i)) \le O(1),$$

where the $O(1)$ term is uniform in i and the last inequality follows from $\|\boldsymbol{\Lambda}_K\|_{\text{op},2} \le \lambda_1 K$, $\|\boldsymbol{\Phi}_{iK}\|_{\text{op},2} \le N_0\sum_{j=1}^{K}\|\phi_j\|_\infty^2$, $\left\|\boldsymbol{\Sigma}_i^{-1}\right\|_{\text{op},2} \le \sigma^{-2}$, $E(\|\mathbf{X}_i - \boldsymbol{\mu}_i\|_2^4\,|\mathbf{T}_i) \le O(1)$ uniformly over i, where the latter is a consequence of the Gaussian process assumption on $X_i(\cdot)$ and $\|\Gamma\|_\infty < \infty$. Thus, $E(\|\tilde{\boldsymbol{\xi}}_{iK}\|_2^2) = O(1)$ uniformly in i which implies $E(n^{-1}\sum_{i=1}^{n}\|\tilde{\boldsymbol{\xi}}_{iK}\|_2^2) = O(1)$ and the second result in (S.124). □

Lemma S6. *Suppose that (X4), (B1)-(B4), (B2)-(B3), (A1)–(A8) hold and consider a sparse design with* $n_i \le N_0 < \infty$*, setting* $a_n = a_{n1}$ *and* $b_n = b_{n1}$*. Let* $\tilde{Z}_i(t) := \sum_{j=1}^{n_i} w_i K_h(T_{ij} - t) \left(\frac{T_{ij}-t}{h}\right)^r (U_{ij}Y_i - C(t))$*, where* $U_{ij} = X(T_{ij}) - \mu(T_{ij})$ *and* $r = 0, 1$*. Then*

$$E[\tilde{Z}_i^2(t)] = O((n^2 h)^{-1}),$$

where the $O((n^2h)^{-1})$ *term is uniform in* i *and* t*.*

Proof of Lemma S6. Observe

$$\begin{aligned}
&E[\tilde{Z}_i^2(t)] \\
&= E\left(\sum_{j=1}^{n_i} w_i^2 K_h^2(T_{ij} - t)\left(\frac{T_{ij}-t}{h}\right)^{2r} (U_{ij}Y_i - C(t))^2\right) \\
&+ E\Big(\sum_{j=1}^{n_i}\sum_{l\neq j} w_i^2 K_h(T_{ij} - t) K_h(T_{il} - t) \\
&\qquad \left(\frac{T_{ij}-t}{h}\right)^{r}\left(\frac{T_{il}-t}{h}\right)^{r} (U_{ij}Y_i - C(t))(U_{il}Y_i - C(t))\Big)
\end{aligned}$$

and note that for any $t_1, t_2 \in \mathcal{T}$, with $\mu_Y = E(Y)$,

$$\begin{aligned}
E(U(t_1)U(t_2)Y^2) &= E(U(t_1)U(t_2)[\mu_Y + \int_{\mathcal{T}} \beta(s)U(s)ds + \epsilon_Y]^2) \\
&= (\mu_Y^2 + \sigma_Y^2)\Gamma(t_1, t_2) + 2\int_{\mathcal{T}} \mu_Y \beta(s) E(U(t_1)U(t_2)U(s))ds \\
&\quad + \int_{\mathcal{T}}\int_{\mathcal{T}} \beta(s_1)\beta(s_2)E(U(t_1)U(t_2)U(s_1)U(s_2))ds_1 ds_2 \\
&= O(1),
\end{aligned}$$

where the $O(1)$ term is uniform over t_1 and t_2, which follows from $\|\Gamma\|_\infty < \infty$ and $U(t) \sim N(0, \Gamma(t,t))$, owing to (X4). This implies that $E((U_{ij}Y_i -$

$C(t))^2|T_{ij})$ is uniformly bounded above, and by a conditioning argument it follows that

$$
\begin{aligned}
&E\left(\sum_{j=1}^{n_i} w_i^2 K_h^2(T_{ij}-t)\left(\frac{T_{ij}-t}{h}\right)^{2r}(U_{ij}Y_i - C(t))^2\right)\\
&\le O(1)E\left(\sum_{j=1}^{n_i} w_i^2 K_h^2(T_{ij}-t)\left(\frac{T_{ij}-t}{h}\right)^{2r}\right)\\
&= O((n^2h)^{-1}),
\end{aligned}
$$

where the last equality is due to $w_i \le n^{-1}$. Let $R_{iqr,h}(t) = w_i K_h(T_{iq} - t)\left(\frac{T_{iq}-t}{h}\right)^r$, $q = j, l$. Since $E((U_{ij}Y_i - C(t))(U_{il}Y_i - C(t))|T_{ij}, T_{il}) = O(1)$ uniformly in i and t, similar arguments as before show that

$$
\begin{aligned}
&E\left(\sum_{j=1}^{n_i}\sum_{l\neq j} R_{ijr,h}(t)R_{ilr,h}(t)(U_{ij}Y_i - C(t))(U_{il}Y_i - C(t))\right)\\
&\le O(1)\sum_{j=1}^{n_i}\sum_{l\neq j} E[R_{ijr,h}(t)]E[R_{ilr,h}(t)]\\
&= O(n^{-2}),
\end{aligned}
$$

whence the result follows. □

Lemma S7. *Suppose that (X4), (B1)-(B4), (B2)-(B3), (A1)–(A8) hold and consider a sparse design with $n_i \le N_0 < \infty$, setting $a_n = a_{n1}$ and $b_n = b_{n1}$. For $r = 0, 1$ we have*

$$
\|\sum_{i=1}^{n}\sum_{j=1}^{n_i} w_i K_h(T_{ij} - \cdot)\left(\frac{T_{ij}-\cdot}{h}\right)^r \epsilon_{ij}Y_i\|_{L^2} = O_p((nh)^{-1/2}), \qquad \text{(S.129)}
$$

and

$$
\|\sum_{i=1}^{n}\sum_{j=1}^{n_i} w_i K_h(T_{ij} - \cdot)\left(\frac{T_{ij}-\cdot}{h}\right)^r (U_{ij}Y_i - C(\cdot))\|_{L^2} = O_p\left(\left(\frac{1}{nh} + h^2\right)^{1/2}\right), \qquad \text{(S.130)}
$$

where $U_{ij} = X(T_{ij}) - \mu(T_{ij})$.

Proof of Lemma S7. Define $Z_i(t) := \sum_{j=1}^{n_i} w_i K_h(T_{ij} - t) \left(\frac{T_{ij}-t}{h}\right)^r \epsilon_{ij} Y_i$. Note that the Z_i are independent and by independence of the ϵ_{ij} along with a conditioning argument, $E(Z_i(t)) = 0$ and

$$E(\|\sum_{i=1}^{n} Z_i\|_{L^2}^2) = \sum_{i=1}^{n} \int_{\mathcal{T}} E(Z_i^2(t))dt,$$

$$\begin{aligned}
E(Z_i^2(t)) &= E\left(\sum_{j=1}^{n_i}\sum_{l=1}^{n_i} w_i^2 K_h(T_{ij}-t)\left(\frac{T_{ij}-t}{h}\right)^r \epsilon_{ij} K_h(T_{il}-t)\left(\frac{T_{il}-t}{h}\right)^r \epsilon_{il} Y_i^2\right) \\
&= \sum_{j=1}^{n_i} E\left(w_i^2 K_h^2(T_{ij}-t)\left(\frac{T_{ij}-t}{h}\right)^{2r} \epsilon_{ij}^2 Y_i^2\right) \\
&= E(Y^2)\sigma^2 \sum_{j=1}^{n_i} E\left(w_i^2 K_h^2(T_{ij}-t)\left(\frac{T_{ij}-t}{h}\right)^{2r}\right) = O((n^2h)^{-1}),
\end{aligned}$$

where the $O(h^{-1})$ is uniform in i and t. Thus $E(\|\sum_{i=1}^{n} Z_i\|_{L^2}^2) = O((nh)^{-1})$ and the first result in (S.129) follows. Defining $\tilde{Z}_i(t) := \sum_{j=1}^{n_i} w_i K_h(T_{ij} - t) \left(\frac{T_{ij}-t}{h}\right)^r (U_{ij}Y_i - C(t))$, we have

$$E(\|\sum_{i=1}^{n} \tilde{Z}_i\|_{L^2}^2) = \sum_{i=1}^{n} \int_{\mathcal{T}} E[\tilde{Z}_i^2(t)]dt + \sum_{i=1}^{n}\sum_{k\neq i} \int_{\mathcal{T}} E(\tilde{Z}_i(t))E(\tilde{Z}_k(t)). \tag{S.131}$$

By a conditioning argument, it follows that

$$\begin{aligned}
|E(\tilde{Z}_i(t))| &= \Big| \sum_{j=1}^{n_i} w_i E\left(K_h(T_{ij}-t)\left(\frac{T_{ij}-t}{h}\right)^r (C(T_{ij})-C(t))\right)\Big| \\
&\leq \sum_{j=1}^{n_i} w_i \int_{-t/h}^{(1-t)/h} |u^r| K(u) |C(t+uh)-C(t)| f(t+uh) du \\
&\leq \sum_{j=1}^{n_i} w_i \sup_{s\in[-1,1]} |C'(s)| \, \|f\|_\infty h \int_{-t/h}^{(1-t)/h} |u^{r+1}| K(u) du \\
&\leq O\left(n^{-1}h\right),
\end{aligned}$$

where the $O\left(n^{-1}h\right)$ is uniform in i and t. This implies $|\sum_{i=1}^n \sum_{k\neq i} \int_{\mathcal{T}} E(\tilde{Z}_i(t))E(\tilde{Z}_k(t))| = O(h^2)$. Combining with (S.131) and Lemma S6, the second result in (S.130) follows. □

Lemma S8. *Suppose that (X4), (B1)-(B4), (A1)–(A8) hold and consider a sparse design with $n_i \leq N_0 < \infty$, setting $a_n = a_{n1}$ and $b_n = b_{n1}$. For $r = 0, 1$,*

$$\|\sum_{i=1}^n \sum_{j=1}^{n_i} w_i K_h(T_{ij}-t)\left(\frac{T_{ij}-t}{h}\right)^r (\mu(T_{ij})-\hat{\mu}(T_{ij}))Y_i\|_{L^2} = O_p(a_n).$$

Proof of Lemma S8. Setting $Z_i := \sum_{j=1}^{n_i} w_i K_h(T_{ij}-t)\left(\frac{T_{ij}-t}{h}\right)^r (\mu(T_{ij})-\hat{\mu}(T_{ij}))Y_i$, note that

$$E\left(\|\sum_{i=1}^n Z_i\|_{L^2}^2\right) = \int_{\mathcal{T}} \sum_{i=1}^n E[Z_i^2(t)]dt + \int_{\mathcal{T}} \sum_{i=1}^n \sum_{k\neq i} E[Z_i(t)Z_k(t)]. \quad \text{(S.132)}$$

Since $|Z_i(t)| \le \|\hat{\mu}-\mu\|_\infty \sum_{j=1}^{n_i} w_i K_h(T_{ij}-t)\left(\frac{|T_{ij}-t|}{h}\right)^r |Y_i|$, it follows that

$$
\begin{aligned}
&E[Z_i^2(t)] \\
&\le E\Big[\|\hat{\mu}-\mu\|_\infty^2 \sum_{j=1}^{n_i}\sum_{l=1}^{n_i} w_i^2 Y_i^2 K_h(T_{ij}-t)K_h(T_{il}-t)\left(\frac{|T_{ij}-t|}{h}\right)^r\left(\frac{|T_{il}-t|}{h}\right)^r\Big] \\
&\le O(a_n^2)\Big\{\sum_{j=1}^{n_i} w_i^2 E(Y^2) E\Big[K_h^2(T_{ij}-t)\left(\frac{T_{ij}-t}{h}\right)^{2r}\Big] \\
&+ \sum_{j=1}^{n_i}\sum_{l\neq j} w_i^2 E(Y^2) E\Big[K_h(T_{ij}-t)\left(\frac{|T_{ij}-t|}{h}\right)^r\Big] E\Big[K_h(T_{il}-t)\left(\frac{|T_{il}-t|}{h}\right)^r\Big]\Big\} \\
&\le O(a_n^2)[O(n^{-2}h^{-1})+O(n^{-2})] \\
&= O(a_n^2 n^{-2}h^{-1}),
\end{aligned} \tag{S.133}
$$

where the first inequality follows from Theorem 5.1 in Zhang and Wang (2016) and the term $O(a_n^2 n^{-2}h^{-1})$ is uniform in i and t. Similarly, for $k \neq i$ and setting $h_{qdr}(t) := \left(\frac{|T_{qd}-t|}{h}\right)^r$, $q = i, k$ and $d = j, l$, we have

$$
\begin{aligned}
&E(|Z_i(t)Z_k(t)|) \\
&\le E\Big[\sum_{j=1}^{n_i}\sum_{l=1}^{n_k} w_i K_h(T_{ij}-t)h_{ijr}(t)|\mu(T_{ij})-\hat{\mu}(T_{ij}) \\
&\qquad\qquad |Y_i w_k K_h(T_{kl}-t)h_{klr}(t)|\mu(T_{kl})-\hat{\mu}(T_{kl})|Y_k\Big] \\
&\le O(a_n^2)\sum_{j=1}^{n_i}\sum_{l=1}^{n_k} w_i w_k E[K_h(T_{ij}-t)h_{ijr}(t)]E[K_h(T_{kl}-t)h_{klr}(t)][E(Y)]^2 \\
&= O(a_n^2 n^{-2}),
\end{aligned}
$$

where the $O(a_n^2 n^{-2})$ term is uniform in i, k and t. Combining this with (S.132) and (S.133) leads to the result. □

Lemma S9. *Suppose that (X4), (B1)-(B4), (A1)–(A8) hold and consider a sparse design with $n_i \leq N_0 < \infty$, setting $a_n = a_{n1}$ and $b_n = b_{n1}$. Then*

$$\|\hat{C} - C\|_{L^2} = O_p\left(\left(\frac{1}{nh} + h^2\right)^{1/2} + a_n\right).$$

Proof of Lemma S9. Proceeding similarly to the proof of Theorem 3.1 in Zhang and Wang (2016), using (S.2),

$$\hat{C}(t) = \frac{S_2(t)\tilde{R}_0(t) - S_1(t)\tilde{R}_1(t)}{S_0(t)S_2(t) - S_1^2(t)},$$

where

$$S_r(t) = \sum_{i=1}^{n}\sum_{j=1}^{n_i} w_i K_h(T_{ij} - t)\left(\frac{T_{ij} - t}{h}\right)^r,$$
$$\tilde{R}_r(t) = \sum_{i=1}^{n}\sum_{j=1}^{n_i} w_i K_h(T_{ij} - t)\left(\frac{T_{ij} - t}{h}\right)^r C_i(T_{ij}),$$

and $r = 0, 1, 2$. Then

$$\hat{C}(t) - C(t) = \frac{(\tilde{R}_0(t) - C(t)S_0(t))S_2(t) - (\tilde{R}_1(t) - C(t)S_1(t))S_1(t)}{S_0(t)S_2(t) - S_1^2(t)}. \tag{S.134}$$

Since $C_i(T_{ij}) = (\tilde{X}_{ij} - \hat{\mu}(T_{ij}))Y_i = (U_{ij} + \epsilon_{ij})Y_i + (\mu(T_{ij}) - \hat{\mu}(T_{ij}))Y_i$, where

$U_{ij} = X(T_{ij}) - \mu(T_{ij}),$

$$
\begin{aligned}
&\|\tilde{R}_0(t) - C(t)S_0(t)\|_{L^2} \\
&\leq \|\sum_{i=1}^{n}\sum_{j=1}^{n_i} w_i K_h(T_{ij}-t)(U_{ij}Y_i - C(t))\|_{L^2} + \|\sum_{i=1}^{n}\sum_{j=1}^{n_i} w_i K_h(T_{ij}-t)\epsilon_{ij}Y_i\|_{L^2} \\
&\quad + \|\sum_{i=1}^{n}\sum_{j=1}^{n_i} w_i K_h(T_{ij}-t)(\mu(T_{ij}) - \hat{\mu}(T_{ij}))Y_i\|_{L^2} \\
&= O_p\left(\left(\frac{1}{nh} + h^2\right)^{1/2}\right) + O_p((nh)^{-1/2}) + O_p(a_n) \\
&= O_p\left(\left(\frac{1}{nh} + h^2\right)^{1/2}\right) + O_p(a_n),
\end{aligned}
$$

where the last equality follows from Lemma S7 and Lemma S8. Similarly

$$
\begin{aligned}
&\|\tilde{R}_1(t) - C(t)S_1(t)\|_{L^2} \\
&\leq \|\sum_{i=1}^{n}\sum_{j=1}^{n_i} w_i K_h(T_{ij}-t)\left(\frac{T_{ij}-t}{h}\right)(U_{ij}Y_i - C(t))\|_{L^2} \\
&\quad + \|\sum_{i=1}^{n}\sum_{j=1}^{n_i} w_i K_h(T_{ij}-t)\left(\frac{T_{ij}-t}{h}\right)\epsilon_{ij}Y_i\|_{L^2} \\
&\quad\quad + \|\sum_{i=1}^{n}\sum_{j=1}^{n_i} w_i K_h(T_{ij}-t)\left(\frac{T_{ij}-t}{h}\right)(\mu(T_{ij}) - \hat{\mu}(T_{ij}))Y_i\|_{L^2} \\
&= O_p\left(\left(\frac{1}{nh} + h^2\right)^{1/2}\right) + O_p(a_n).
\end{aligned}
$$

These along with (S.134) and similar arguments as in the proof of Theorem 4.1 in Zhang and Wang (2016) show that $S_0(t)S_2(t) - S_1^2(t)$ is positive and bounded away from 0 with probability tending to 1 and $\sup_{t\in\mathcal{T}}|S_r(t)| = O_p(1)$, $r = 1, 2$. The result then follows. $\square$

Recall that the eigenpairs of the integral operator $\hat{\Xi}$ associated with $\hat{\Gamma}$ are

$(\hat{\lambda}_k, \hat{\phi}_k)$, and those of Ξ are (λ_k, ϕ_k), $k \geq 1$.

Lemma S10. *Suppose that (X4), (B1)-(B4), (A1)–(A8) hold and consider a sparse design with $n_i \leq N_0 < \infty$, setting $a_n = a_{n1}$ and $b_n = b_{n1}$. Then, setting $\tau_M = \sum_{m=1}^{M} \frac{1}{\lambda_m}$, for large enough n, the following relations hold almost surely,*

$$\sum_{m=1}^{M} \frac{|\hat{\sigma}_m - \sigma_m|}{\lambda_m} = \tau_M \|\hat{C} - C\|_{L^2} + \tau_M^{1/2} O(c_n^{\rho}), \tag{S.135}$$

$$\sum_{m=1}^{M} |\hat{\sigma}_m - \sigma_m| \frac{|\hat{\lambda}_m - \lambda_m|}{|\hat{\lambda}_m| \lambda_m} \leq O(c_n^{2\rho}) + \|\hat{C} - C\|_{L^2} \tau_M^{1/2} O(c_n^{\rho}), \tag{S.136}$$

$$\sum_{m=1}^{M} |\sigma_m| \frac{|\hat{\lambda}_m - \lambda_m|}{|\hat{\lambda}_m| \lambda_m} \leq O(c_n) \tau_M, \tag{S.137}$$

$$\sum_{m=1}^{M} \left| \frac{\hat{\sigma}_m}{\hat{\lambda}_m} - \frac{\sigma_m}{\lambda_m} \right| \|\hat{\phi}_m - \phi_m\|_{L^2} \leq O(c_n^{2\rho}) + O(c_n^{\rho})(\|\hat{C} - C\|_{L^2} + c_n) \tau_M^{1/2}, \tag{S.138}$$

$$\sum_{m=1}^{M} \frac{|\sigma_m|}{\lambda_m} \|\hat{\phi}_m - \phi_m\|_{L^2} \leq O(c_n) \upsilon_M, \tag{S.139}$$

Proof of Lemma S10. First note

$$\begin{aligned} \sum_{m=1}^{M} \frac{1}{\delta_m} &\leq \left(\sum_{m=1}^{M} \frac{1}{\lambda_m \delta_m^2} \right)^{1/2} \left(\sum_{m=1}^{M} \lambda_m \right)^{1/2} \\ &\leq \left(\sum_{m=1}^{M} \frac{1}{\sqrt{\lambda_m} \delta_m} \right) \left(\sum_{m=1}^{\infty} \lambda_m \right)^{1/2} \\ &= O(c_n^{\rho - 1}), \end{aligned}$$

implying $c_n \upsilon_M = O(c_n^{\rho}) = o(1)$ as $n \to \infty$. By the Cauchy–Schwarz inequality and from Theorem 5.2 in Zhang and Wang (2016), we have $\|\hat{\Xi} - \Xi\|_{\text{op}} = O(a_n +$

b_n) a.s.. Note that from the orthonormality of the ϕ_k and using perturbation results (Bosq, 2000), we have $\|\hat{\phi}_k - \phi_k\|_{L^2} \leq 2\sqrt{2}\|\hat{\Xi} - \Xi\|_{\text{op}}/\delta_k$, $k \geq 1$, so that for any $m \geq 1$

$$|\hat{\sigma}_m - \sigma_m| = |\langle \hat{C}, \hat{\phi}_m \rangle_{L^2} - \langle C, \phi_m \rangle_{L^2}| \leq 2\sqrt{2}\|\hat{C} - C\|_{L^2} \frac{\|\hat{\Xi} - \Xi\|_{\text{op}}}{\delta_m} + \|\hat{C} - C\|_{L^2} + 2\sqrt{2}\|C\|_{L^2} \frac{\|\hat{\Xi} - \Xi\|_{\text{op}}}{\delta_m}, \tag{S.140}$$

and from $\delta_m \leq \lambda_m$,

$$\sum_{m=1}^{M} \frac{\|\hat{\Xi} - \Xi\|_{\text{op}}}{\lambda_m \delta_m} \leq \tau_M^{1/2} \sum_{m=1}^{M} \frac{\|\hat{\Xi} - \Xi\|_{\text{op}}}{\sqrt{\lambda_m}\delta_m} = \tau_M^{1/2} O(c_n^{\rho}) \quad \text{a.s..} \tag{S.141}$$

Thus

$$\sum_{m=1}^{M} \frac{|\hat{\sigma}_m - \sigma_m|}{\lambda_m} \leq \tau_M^{1/2} O(c_n^{\rho}) \|\hat{C} - C\|_{L^2} + \tau_M \|\hat{C} - C\|_{L^2} + \tau_M^{1/2} O(c_n^{\rho}) \quad \text{a.s.}$$
$$= \tau_M \|\hat{C} - C\|_{L^2} + \tau_M^{1/2} O(c_n^{\rho}),$$

which shows the first result in (S.135). Since $M = M(n)$ is such that $\sum_{m=1}^{M} \frac{1}{\sqrt{\lambda_m}\delta_m} = O(c_n^{\rho-1})$ as $n \to \infty$, then $\sum_{m=1}^{M} \|\hat{\Xi} - \Xi\|_{\text{op}} \lambda_m^{-1/2} \delta_m^{-1} = O(c_n^{\rho}) = o(1)$ a.s. and $\lambda_M = o(1)$ as $n \to \infty$. Thus, for large enough n we have $\lambda_M < 1$ and $\|\hat{\Xi} - \Xi\|_{\text{op}} \lambda_M^{-1/2} \delta_M^{-1} \leq \sum_{m=1}^{M} \|\hat{\Xi} - \Xi\|_{\text{op}} \lambda_m^{-1/2} \delta_m^{-1} \leq 1/2$ a.s., so that $\|\hat{\Xi} - \Xi\|_{\text{op}} \leq \lambda_M^{1/2} \delta_M / 2 \leq \delta_M / 2 \leq \lambda_M / 2$ a.s.. This shows that there exists $n_0 \geq 1$ such that for all $n \geq n_0$ it holds that $\|\hat{\Xi} - \Xi\|_{\text{op}} \leq \lambda_M / 2$ a.s.. Then $|\hat{\lambda}_m - \lambda_m| \leq \|\hat{\Xi} - \Xi\|_{\text{op}}$

implies $|\hat{\lambda}_m| \geq \lambda_m/2$ a.s. for large enough n. With (S.140), (S.141),

$$\begin{aligned}\sum_{m=1}^{M}|\hat{\sigma}_m-\sigma_m|\frac{|\hat{\lambda}_m-\lambda_m|}{|\hat{\lambda}_m|\lambda_m} &\leq 2\sum_{m=1}^{M}|\hat{\sigma}_m-\sigma_m|\frac{\|\hat{\Xi}-\Xi\|_{\text{op}}}{\lambda_m^2}\\ &\leq 4\sqrt{2}\|\hat{C}-C\|_{L^2}\sum_{m=1}^{M}\frac{\|\hat{\Xi}-\Xi\|_{\text{op}}^2}{\lambda_m^2\delta_m}+2\|\hat{C}-C\|_{L^2}\sum_{m=1}^{M}\frac{\|\hat{\Xi}-\Xi\|_{\text{op}}}{\lambda_m^2}\\ &\quad+4\sqrt{2}\|C\|_{L^2}\sum_{m=1}^{M}\frac{\|\hat{\Xi}-\Xi\|_{\text{op}}^2}{\lambda_m^2\delta_m}\\ &\leq \|\hat{C}-C\|_{L^2}O(c_n^{2\rho})+\|\hat{C}-C\|_{L^2}\tau_M^{1/2}O(c_n^{\rho})+O(c_n^{2\rho}) \quad \text{a.s.}\\ &= O(c_n^{2\rho})+\|\hat{C}-C\|_{L^2}\tau_M^{1/2}O(c_n^{\rho}) \quad \text{a.s.,}\end{aligned}$$

for large enough n, implying the second result in (S.136). Similarly, for large enough n and a.s.

$$\sum_{m=1}^{M}|\sigma_m|\frac{|\hat{\lambda}_m-\lambda_m|}{|\hat{\lambda}_m|\lambda_m} \leq 2\sum_{m=1}^{M}|\sigma_m|\frac{|\hat{\lambda}_m-\lambda_m|}{\lambda_m^2} \leq O(c_n)\left(\sum_{m=1}^{M}\frac{\sigma_m^2}{\lambda_m^2}\right)^{1/2}\tau_M = O(c_n)\tau_M,$$

where the last equality is due to $\sum_{m=1}^{\infty}\sigma_m^2/\lambda_m^2 < \infty$. This shows the third result in (S.137). Now,

$$\begin{aligned}&\sum_{m=1}^{M}\left|\frac{\hat{\sigma}_m}{\hat{\lambda}_m}-\frac{\sigma_m}{\lambda_m}\right|\|\hat{\phi}_m-\phi_m\|_{L^2}\\ &\leq \sum_{m=1}^{M}\frac{|\hat{\sigma}_m-\sigma_m|}{\lambda_m}\|\hat{\phi}_m-\phi_m\|_{L^2}+\sum_{m=1}^{M}\frac{|\sigma_m||\hat{\lambda}_m-\lambda_m|}{|\hat{\lambda}_m|\lambda_m}\|\hat{\phi}_m-\phi_m\|_{L^2}\\ &\quad+\sum_{m=1}^{M}\frac{|\hat{\sigma}_m-\sigma_m||\hat{\lambda}_m-\lambda_m|}{|\hat{\lambda}_m|\lambda_m}\|\hat{\phi}_m-\phi_m\|_{L^2}. \qquad \text{(S.142)}\end{aligned}$$

From (S.140), (S.141) and using that $\|\hat{\phi}_m-\phi_m\|_{L^2} \leq 2\sqrt{2}\|\hat{\Xi}-\Xi\|_{\text{op}}/\delta_m$, we obtain

$$
\begin{aligned}
&\sum_{m=1}^{M} \frac{|\hat{\sigma}_m - \sigma_m|}{\lambda_m} \|\hat{\phi}_m - \phi_m\|_{L^2} \\
&\leq 8\|\hat{C} - C\|_{L^2} \sum_{m=1}^{M} \|\hat{\Xi} - \Xi\|_{\mathrm{op}}^2 \frac{1}{\lambda_m \delta_m^2} + 2\sqrt{2}\|\hat{C} - C\|_{L^2} \sum_{m=1}^{M} \frac{\|\hat{\Xi} - \Xi\|_{\mathrm{op}}}{\lambda_m \delta_m} \\
&\quad + 8\|C\|_{L^2} \sum_{m=1}^{M} \frac{\|\hat{\Xi} - \Xi\|_{\mathrm{op}}^2}{\lambda_m \delta_m^2} \\
&\leq \|\hat{C} - C\|_{L^2} O(c_n^{2\rho}) + \|\hat{C} - C\|_{L^2} \tau_M^{1/2} O(c_n^{\rho}) + O(c_n^{2\rho}) \quad \text{a.s.} \\
&= O(c_n^{2\rho}) + \|\hat{C} - C\|_{L^2} \tau_M^{1/2} O(c_n^{\rho}) \quad \text{a.s..}
\end{aligned} \tag{S.143}
$$

For large enough n,

$$
\begin{aligned}
\sum_{m=1}^{M} \frac{|\sigma_m||\hat{\lambda}_m - \lambda_m|}{|\hat{\lambda}_m|\lambda_m} \|\hat{\phi}_m - \phi_m\|_{L^2} &\leq 4\sqrt{2} \sum_{m=1}^{M} |\sigma_m| \frac{\|\hat{\Xi} - \Xi\|_{\mathrm{op}}^2}{\lambda_m^2 \delta_m} \quad \text{a.s.} \\
&\leq \left(\sum_{m=1}^{M} \frac{\sigma_m^2}{\lambda_m^2} \right)^{1/2} O(c_n^{1+\rho}) \tau_M^{1/2} \quad \text{a.s.} \\
&= O(c_n^{1+\rho}) \tau_M^{1/2}.
\end{aligned} \tag{S.144}
$$

Similarly, from (S.140) we obtain

$$
\begin{aligned}
&\sum_{m=1}^{M} \frac{|\hat{\sigma}_m - \sigma_m||\hat{\lambda}_m - \lambda_m|}{|\hat{\lambda}_m|\lambda_m} \|\hat{\phi}_m - \phi_m\|_{L^2} \\
&\leq 4\sqrt{2} \sum_{m=1}^{M} |\hat{\sigma}_m - \sigma_m| \frac{\|\hat{\Xi} - \Xi\|_{\text{op}}^2}{\lambda_m^2 \delta_m} \\
&\leq 16\|\hat{C} - C\|_{L^2} \sum_{m=1}^{M} \frac{\|\hat{\Xi} - \Xi\|_{\text{op}}^3}{\lambda_m^2 \delta_m^2} + 4\sqrt{2}\|\hat{C} - C\|_{L^2} \sum_{m=1}^{M} \frac{\|\hat{\Xi} - \Xi\|_{\text{op}}^2}{\lambda_m^2 \delta_m} \\
&\quad + 16\|C\|_{L^2} \sum_{m=1}^{M} \frac{\|\hat{\Xi} - \Xi\|_{\text{op}}^3}{\lambda_m^2 \delta_m^2} \\
&\leq O(c_n^{1+2\rho})\tau_M \|\hat{C} - C\|_{L^2} + O(c_n^{2\rho})\|\hat{C} - C\|_{L^2} + O(c_n^{1+2\rho})\tau_M \quad \text{a.s.} \\
&= O(c_n^{1+2\rho})\tau_M + O(c_n^{2\rho})\|\hat{C} - C\|_{L^2} \quad \text{a.s..} \qquad \text{(S.145)}
\end{aligned}
$$

Combining (S.142), (S.143), (S.144) and (S.145) with the fact that $c_n \tau_M \leq c_n \upsilon_M = o(1)$ as $n \to \infty$, which was already shown, leads to the fourth result in (S.138). Finally

$$
\begin{aligned}
&\sum_{m=1}^{M} \frac{|\sigma_m|}{\lambda_m} \|\hat{\phi}_m - \phi_m\|_{L^2} \\
&\leq 2\sqrt{2} \sum_{m=1}^{M} \frac{|\sigma_m| \, \|\hat{\Xi} - \Xi\|_{\text{op}}}{\lambda_m \delta_m} \\
&\leq \left(\sum_{m=1}^{M} \frac{\sigma_m^2}{\lambda_m^2} \right)^{1/2} \|\hat{\Xi} - \Xi\|_{\text{op}} \upsilon_M = O(c_n)\upsilon_M \quad \text{a.s.,}
\end{aligned}
$$

which shows the last result in (S.139). □

The next lemma provides the L^2 convergence of the empirical estimate $\hat{\beta}_M$ towards β, which is required to construct the estimated predictive distribution

$\hat{\mathcal{P}}_{iK}$. Recall that

$$\hat{\beta}_M(t) := \sum_{m=1}^{M} \frac{\hat{\sigma}_m}{\hat{\lambda}_m}\hat{\phi}_m(t), \quad t \in \mathcal{T},$$

$\Theta_M = \Big\| \sum_{m \geq M+1} \frac{\sigma_m}{\lambda_m}\phi_m \Big\|_{L^2}$ and $\tau_M = \sum_{m=1}^{M} \lambda_m^{-1}$.

Lemma S11. *Suppose that (X4), (B1)-(B4), (A1)–(A8) hold and consider a sparse design with $n_i \leq N_0 < \infty$, setting $a_n = a_{n1}$ and $b_n = b_{n1}$. Let $K \geq 1$. Then*

$$\|\hat{\beta}_M - \beta\|_{L^2} = O_p(r_n), \tag{S.146}$$

and

$$\int_{\mathcal{T}} \hat{\beta}_M(t)\hat{\phi}_k(t)dt = \int_{\mathcal{T}} \beta(t)\phi_k(t)dt + O_p(r_n), \tag{S.147}$$

where $r_n = c_n \upsilon_M + c_n^{\rho}\tau_M^{1/2} + \tau_M\Big[\big(\frac{1}{nh} + h^2\big)^{1/2} + a_n\Big] + \Theta_M$ and $k = 1, \ldots, K$.

Proof of Lemma S11. Observe

$$\|\hat{\beta}_M - \beta\|_{L^2} \leq \sum_{m=1}^{M} \Big\| \frac{\hat{\sigma}_m}{\hat{\lambda}_m}\hat{\phi}_m - \frac{\sigma_m}{\lambda_m}\phi_m \Big\|_{L^2} + \Big\| \sum_{m \geq M+1} \frac{\sigma_m}{\lambda_m}\phi_m \Big\|_{L^2}, \tag{S.148}$$

and

$$\sum_{m=1}^{M} \Big\| \frac{\hat{\sigma}_m}{\hat{\lambda}_m}\hat{\phi}_m - \frac{\sigma_m}{\lambda_m}\phi_m \Big\|_{L^2} \leq \sum_{m=1}^{M} \Big| \frac{\hat{\sigma}_m}{\hat{\lambda}_m} - \frac{\sigma_m}{\lambda_m} \Big| \|\hat{\phi}_m - \phi_m\|_{L^2} + \sum_{m=1}^{M} \Big| \frac{\hat{\sigma}_m}{\hat{\lambda}_m} - \frac{\sigma_m}{\lambda_m} \Big| + \sum_{m=1}^{M} \frac{|\sigma_m|}{\lambda_m} \|\hat{\phi}_m - \phi_m\|_{L^2}. \tag{S.149}$$

By the triangle inequality and Lemma S10, we have that for large enough n

$$\begin{aligned}
\sum_{m=1}^{M} \Big| \frac{\hat{\sigma}_m}{\hat{\lambda}_m} - \frac{\sigma_m}{\lambda_m} \Big| &\leq \sum_{m=1}^{M} \frac{|\hat{\sigma}_m - \sigma_m|}{\lambda_m} + \sum_{m=1}^{M} \frac{|\hat{\sigma}_m - \sigma_m||\hat{\lambda}_m - \lambda_m|}{|\hat{\lambda}_m|\lambda_m} + \sum_{m=1}^{M} \frac{|\sigma_m||\hat{\lambda}_m - \lambda_m|}{|\hat{\lambda}_m|\lambda_m} \\
&= \tau_M \|\hat{C} - C\|_{L^2} + O(c_n^{\rho})\tau_M^{1/2} + O(c_n)\tau_M \quad \text{a.s.} \\
&= \tau_M \|\hat{C} - C\|_{L^2} + O(c_n^{\rho})\tau_M^{1/2} \quad \text{a.s.},
\end{aligned}$$

where the second equality is due to $c_n\tau_M = c_n^\rho \tau_M^{1/2} c_n^{1-\rho}\tau_M^{1/2} = o(1)c_n^\rho \tau_M^{1/2}$, and

$$\sum_{m=1}^{M}\Big|\frac{\hat\sigma_m}{\hat\lambda_m}-\frac{\sigma_m}{\lambda_m}\Big|\|\hat\phi_m-\phi_m\|_{L^2}+\sum_{m=1}^{M}\frac{|\sigma_m|}{\lambda_m}\|\hat\phi_m-\phi_m\|_{L^2}$$
$$\le O(c_n^{2\rho})+O(c_n^\rho)\|\hat C - C\|_{L^2}\tau_M^{1/2}+O(c_n)\upsilon_M.$$

With (S.148), (S.149) and the fact that $\upsilon_M = O(c_n^{\rho-1})$ as $n\to\infty$, which was shown in the proof of Lemma S10, we arrive at

$$\|\hat\beta_M-\beta\|_{L^2}\le O(c_n)\upsilon_M+O(c_n^\rho)\tau_M^{1/2}+\tau_M\|\hat C - C\|_{L^2}+\Big\|\sum_{m\ge M+1}\frac{\sigma_m}{\lambda_m}\phi_m\Big\|_{L^2}$$

and the result in (S.146) follows from Lemma S9. Finally, recalling that $\hat\beta_k = \int_{\mathcal{T}}\hat\beta_M(t)\hat\phi_k(t)dt$ and $\beta_k = \int_{\mathcal{T}}\beta(t)\phi_k(t)dt$, we have

$$\begin{aligned}|\hat\beta_k-\beta_k| &= |\int_{\mathcal{T}}[\hat\beta_M(t)\hat\phi_k(t)-\beta(t)\phi_k(t)]dt|\\ &\le \|\hat\beta_M-\beta\|_{L^2}\|\hat\phi_k-\phi_k\|_{L^2}+\|\hat\beta_M-\beta\|_{L^2}+\|\beta\|_{L^2}\|\hat\phi_k-\phi_k\|_{L^2}\\ &= O_p(r_n+a_n+b_n)=O_p(r_n),\end{aligned}$$

where the second equality is due to the fact that $\|\hat\phi_k-\phi_k\|_{L^2}\le O(a_n+b_n)$ a.s., which follows from the proof of Lemma S10. This shows the second result in (S.147). □

We remark that in the sparse case when choosing the optimal bandwidth $h\asymp n^{-1/3}$, then the rate

$$\tau_M[\big((nh)^{-1}+h^2\big)^{1/2}+a_n],$$

is faster than $c_n\upsilon_M$ and thus the rate r_n is equivalent to α_n defined as in Theorem 5. Recall that $\mathcal{P}_{iK}$ corresponds to the true predictive distribution $\eta_{iK}|\mathbf{X}_i,\mathbf{T}_i$, or equivalently $N(\beta_0+\boldsymbol{\beta}_K^T\tilde{\boldsymbol{\xi}}_{iK}, \boldsymbol{\beta}_K^T\boldsymbol{\Sigma}_{iK}\boldsymbol{\beta}_K)$, while $\tilde{\mathcal{P}}_{iK} \overset{d}{=} N(\beta_0+\boldsymbol{\beta}_K^T\hat{\boldsymbol{\xi}}_{iK}, \boldsymbol{\beta}_K^T\hat{\boldsymbol{\Sigma}}_{iK}\boldsymbol{\beta}_K)$ corresponds to an intermediate target, replacing population quantities by their estimated counterparts but keeping the true intercept and slope coefficients β_0 and $\boldsymbol{\beta}_K$. Also $\hat{\mathcal{P}}_{iK}$ corresponds to the estimated predictive distribution, i.e. $\hat{\mathcal{P}}_{iK} \overset{d}{=} N(\hat{\beta}_0 + \hat{\boldsymbol{\beta}}_K^T\hat{\boldsymbol{\xi}}_{iK}, \hat{\boldsymbol{\beta}}_K^T\hat{\boldsymbol{\Sigma}}_{iK}\hat{\boldsymbol{\beta}}_K)$. Finally, recall that $F_{iK}(t), \tilde{F}_{iK}(t)$ and $\hat{F}_{iK}$ are the distribution functions associated with $\mathcal{P}_{iK}, \tilde{\mathcal{P}}_{iK}$ and $\hat{\mathcal{P}}_{iK}$, respectively. We require the following auxiliary lemma.

Lemma S12. *Under the conditions of Theorem 5, it holds that*

$$\|\boldsymbol{\Sigma}_{iK} - \hat{\boldsymbol{\Sigma}}_{iK}\|_F = O(N_0^{5/2}(a_n + b_n)),$$

a.s. as $n \to \infty$.

Proof of Lemma S12. Note that

$$\begin{aligned}\boldsymbol{\Sigma}_{iK} - \hat{\boldsymbol{\Sigma}}_{iK} &= (\boldsymbol{\Lambda}_K - \hat{\boldsymbol{\Lambda}}_K) + \hat{\boldsymbol{\Lambda}}_K\hat{\boldsymbol{\Phi}}_{iK}^T\hat{\boldsymbol{\Sigma}}_i^{-1}\hat{\boldsymbol{\Phi}}_{iK}\hat{\boldsymbol{\Lambda}}_K - \boldsymbol{\Lambda}_K\boldsymbol{\Phi}_{iK}^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Phi}_{iK}\boldsymbol{\Lambda}_K \\ &= (\boldsymbol{\Lambda}_K - \hat{\boldsymbol{\Lambda}}_K) + (\hat{\boldsymbol{\Lambda}}_K\hat{\boldsymbol{\Phi}}_{iK}^T - \boldsymbol{\Lambda}_K\boldsymbol{\Phi}_{iK}^T)\hat{\boldsymbol{\Sigma}}_i^{-1}\hat{\boldsymbol{\Phi}}_{iK}\hat{\boldsymbol{\Lambda}}_K \\ &\quad + \boldsymbol{\Lambda}_K\boldsymbol{\Phi}_{iK}^T(\hat{\boldsymbol{\Sigma}}_i^{-1}\hat{\boldsymbol{\Phi}}_{iK}\hat{\boldsymbol{\Lambda}}_K - \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Phi}_{iK}\boldsymbol{\Lambda}_K). \end{aligned} \tag{S.150}$$

Denoting by $C_i := (\hat{\boldsymbol{\Sigma}}_i^{-1}\hat{\boldsymbol{\Phi}}_{iK}\hat{\boldsymbol{\Lambda}}_K - \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Phi}_{iK}\boldsymbol{\Lambda}_K)$, we have

$$C_i = (\hat{\boldsymbol{\Sigma}}_i^{-1} - \boldsymbol{\Sigma}_i^{-1})(\hat{\boldsymbol{\Phi}}_{iK}\hat{\boldsymbol{\Lambda}}_K - \boldsymbol{\Phi}_{iK}\boldsymbol{\Lambda}_K) + \boldsymbol{\Sigma}_i^{-1}(\hat{\boldsymbol{\Phi}}_{iK}\hat{\boldsymbol{\Lambda}}_K - \boldsymbol{\Phi}_{iK}\boldsymbol{\Lambda}_K) + (\hat{\boldsymbol{\Sigma}}_i^{-1} - \boldsymbol{\Sigma}_i^{-1})\boldsymbol{\Phi}_{iK}\boldsymbol{\Lambda}_K, \tag{S.151}$$

where

$$\hat{\boldsymbol{\Phi}}_{iK}\hat{\boldsymbol{\Lambda}}_K - \boldsymbol{\Phi}_{iK}\boldsymbol{\Lambda}_K = (\hat{\boldsymbol{\Phi}}_{iK} - \boldsymbol{\Phi}_{iK})(\hat{\boldsymbol{\Lambda}}_K - \boldsymbol{\Lambda}_K) + \boldsymbol{\Phi}_{iK}(\hat{\boldsymbol{\Lambda}}_K - \boldsymbol{\Lambda}_K) + (\hat{\boldsymbol{\Phi}}_{iK} - \boldsymbol{\Phi}_{iK})\boldsymbol{\Lambda}_K. \tag{S.152}$$

Note that $\|\hat{\mathbf{\Phi}}_{iK} - \mathbf{\Phi}_{iK}\|_F \le \sqrt{N_0 K} \max_{1\le k\le K}\|\hat{\phi}_k - \phi_k\|_\infty = O(\sqrt{N_0}(a_n + b_n))$ a.s. as $n \to \infty$, which follows similarly as in Proposition 1 in Dai *et al.* (2018) by employing Theorem 5.1 and 5.2 in Zhang and Wang (2016). Using perturbation results (Bosq, 2000), Theorem 5.2 in Zhang and Wang (2016) and the Cauchy Schwarz inequality, it follows that $|\hat{\lambda}_k - \lambda_k| \le \|\Gamma - \hat{\Gamma}\|_\infty = O(a_n + b_n)$ a.s. as $n \to \infty$. Thus $\|\hat{\mathbf{\Lambda}}_K - \mathbf{\Lambda}_K\|_F \le \sqrt{K} \max_{1\le k\le K}\|\hat{\lambda}_k - \lambda_k\|_\infty = O(a_n + b_n)$ a.s. as $n \to \infty$. Furthermore, from the proof of Theorem 2 in Dai *et al.* (2018) we have $\|\hat{\mathbf{\Sigma}}_i^{-1} - \mathbf{\Sigma}_i^{-1}\|_{\text{op},2} = O(N_0(a_n + b_n))$ a.s. which implies $\|\hat{\mathbf{\Sigma}}_i^{-1} - \mathbf{\Sigma}_i^{-1}\|_F \le \sqrt{N_0}\|\hat{\mathbf{\Sigma}}_i^{-1} - \mathbf{\Sigma}_i^{-1}\|_{\text{op},2} = O(N_0^{3/2}(a_n + b_n))$ a.s. as $n \to \infty$. Thus, from (S.151) and (S.152), $\|\mathbf{\Sigma}_i^{-1}\|_{\text{op},2} \le \sigma^{-2}$ and $\|\hat{\mathbf{\Phi}}_{iK}\|_F \le \sqrt{N_0 K} \max_{1\le k\le K}\|\phi_k\|_\infty$, it follows that $\|\hat{\mathbf{\Phi}}_{iK}\hat{\mathbf{\Lambda}}_K - \mathbf{\Phi}_{iK}\mathbf{\Lambda}_K\|_F = O(\sqrt{N_0}(a_n + b_n))$ and $\|C_i\|_F = O(N_0^2(a_n + b_n))$ a.s. as $n \to \infty$. From (S.150) and using that

$$\begin{aligned} \|(\hat{\mathbf{\Lambda}}_K\hat{\mathbf{\Phi}}_{iK}^T - \mathbf{\Lambda}_K\mathbf{\Phi}_{iK}^T)\hat{\mathbf{\Sigma}}_i^{-1}\hat{\mathbf{\Phi}}_{iK}\hat{\mathbf{\Lambda}}_K\|_F &= \|(\hat{\mathbf{\Lambda}}_K\hat{\mathbf{\Phi}}_{iK}^T - \mathbf{\Lambda}_K\mathbf{\Phi}_{iK}^T)\left(C_i + \mathbf{\Sigma}_i^{-1}\mathbf{\Phi}_{iK}\mathbf{\Lambda}_K\right)\|_F \\ &= O(N_0(a_n + b_n)) + O(N_0^{5/2}(a_n + b_n)^2) \text{ a.s.,} \end{aligned} \tag{S.153}$$

as $n \to \infty$, we obtain $\|\mathbf{\Sigma}_{iK} - \hat{\mathbf{\Sigma}}_{iK}\|_F = O(N_0^{5/2}(a_n + b_n))$ a.s. as $n \to \infty$, which shows the result. □

The following auxiliary lemmas will be used in the proof of Theorem 6.

Lemma S13. *Suppose that (X4), (B1)-(B4), (A1)–(A8) hold and consider a sparse design with $n_i \le N_0 < \infty$, setting $a_n = a_{n1}$ and $b_n = b_{n1}$. Then*

$$n^{-1}\sum_{i=1}^{n}(\tilde{\eta}_{iK} - \hat{\eta}_{iK})\epsilon_{iY} = O_p(\alpha_n),$$

where $\hat{\eta}_{iK} = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_K^T \hat{\boldsymbol{\xi}}_{iK}$, and $(\hat{\beta}_0, \hat{\boldsymbol{\beta}}_K^T)^T$ are the estimates in the functional linear model as in Theorem 5.

Proof of Lemma S13. By the Cauchy–Schwarz inequality

$$|n^{-1}\sum_{i=1}^{n}(\tilde{\eta}_{iK} - \hat{\eta}_{iK})\epsilon_{iY}| \leq \left(n^{-1}\sum_{i=1}^{n}(\tilde{\eta}_{iK} - \hat{\eta}_{iK})^2\right)^{1/2}\left(n^{-1}\sum_{i=1}^{n}\epsilon_{iY}^2\right)^{1/2}, \tag{S.154}$$

where $(n^{-1}\sum_{i=1}^{n}\epsilon_{iY}^2)^{1/2} = O_p(1)$, whence $|\tilde{\eta}_{iK} - \hat{\eta}_{iK}| \leq |\beta_0 - \hat{\beta}_0| + \|\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K\|_2\|\tilde{\boldsymbol{\xi}}_{iK}\|_2 + \|\hat{\boldsymbol{\beta}}_K\|_2\|\tilde{\boldsymbol{\xi}}_{iK} - \hat{\boldsymbol{\xi}}_{iK}\|_2$, and then

$$\begin{aligned}(\tilde{\eta}_{iK} - \hat{\eta}_{iK})^2 \leq{}& (\beta_0 - \hat{\beta}_0)^2 + \|\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K\|_2^2\|\tilde{\boldsymbol{\xi}}_{iK}\|_2^2 + \|\hat{\boldsymbol{\beta}}_K\|_2^2\|\tilde{\boldsymbol{\xi}}_{iK} - \hat{\boldsymbol{\xi}}_{iK}\|_2^2\\ &+ 2|\beta_0 - \hat{\beta}_0|\|\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K\|_2\|\tilde{\boldsymbol{\xi}}_{iK}\|_2 + 2|\beta_0 - \hat{\beta}_0|\|\hat{\boldsymbol{\beta}}_K\|_2\|\tilde{\boldsymbol{\xi}}_{iK} - \hat{\boldsymbol{\xi}}_{iK}\|_2\\ &+ 2\|\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K\|_2\|\tilde{\boldsymbol{\xi}}_{iK}\|_2\|\hat{\boldsymbol{\beta}}_K\|_2\|\tilde{\boldsymbol{\xi}}_{iK} - \hat{\boldsymbol{\xi}}_{iK}\|_2.\end{aligned}$$

From Lemma S11 we have $|\beta_0 - \hat{\beta}_0| = O_p(n^{-1/2})$ and $\|\hat{\boldsymbol{\beta}}_K - \boldsymbol{\beta}_K\|_2 = O_p(\alpha_n)$, which combined with Lemma S5 and the Cauchy–Schwarz inequality leads to

$$n^{-1}\sum_{i=1}^{n}(\tilde{\eta}_{iK} - \hat{\eta}_{iK})^2 = O_p((\alpha_n)^2). \tag{S.155}$$

The result then follows from (S.154) and (S.155). □

Lemma S14. *Under the conditions of Theorem 6, it holds that*

$$n^{-1}\sum_{i=1}^{n}(\eta_{iK} - \tilde{\eta}_{iK})^2 - \boldsymbol{\beta}_K^T E(\boldsymbol{\Sigma}_{1K})\boldsymbol{\beta}_K = O_p(n^{-1/2}).$$

Proof of Lemma S14. Since $\boldsymbol{\xi}_{iK} - \tilde{\boldsymbol{\xi}}_{iK}|\mathbf{T}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{iK})$, by conditioning on $\mathbf{T}_i$,

$$E(\eta_{iK} - \tilde{\eta}_{iK})^2 = E\left(E\left[\left(\boldsymbol{\beta}_K^T(\boldsymbol{\xi}_{iK} - \tilde{\boldsymbol{\xi}}_{iK})\right)^2\Big|\mathbf{T}_i\right]\right) = \boldsymbol{\beta}_K^T E(\boldsymbol{\Sigma}_{1K})\boldsymbol{\beta}_K,$$

where the last equality is due to the fact that $n_i = m_0$ implies that $\mathbf{\Sigma}_{iK}$ are a sequence of i.i.d. random positive definite matrices. Similarly, since $\eta_{iK} - \tilde{\eta}_{iK}|\mathbf{T}_i \sim N(0, \boldsymbol{\beta}_K^T \mathbf{\Sigma}_{iK} \boldsymbol{\beta}_K)$ we have $E((\eta_{iK} - \tilde{\eta}_{iK})^4|\mathbf{T}_i) = 3(\boldsymbol{\beta}_K^T \mathbf{\Sigma}_{iK} \boldsymbol{\beta}_K)^2$ and thus

$$\begin{aligned}
\text{Var}((\eta_{iK} - \tilde{\eta}_{iK})^2) &= E(\text{ Var}((\eta_{iK} - \tilde{\eta}_{iK})^2|\mathbf{T}_i)) + \text{ Var}(\boldsymbol{\beta}_K^T \mathbf{\Sigma}_{iK} \boldsymbol{\beta}_K) \\
&= 2E((\boldsymbol{\beta}_K^T \mathbf{\Sigma}_{iK} \boldsymbol{\beta}_K)^2) + \text{ Var}(\boldsymbol{\beta}_K^T \mathbf{\Sigma}_{iK} \boldsymbol{\beta}_K) \\
&= O(1),
\end{aligned}$$

where the $O(1)$ term is uniform in i since $\|\mathbf{\Sigma}_{iK}\|_{\text{op}}$ is uniformly bounded in the sparse case. Since the $\eta_{iK} - \tilde{\eta}_{iK}$ are independent, the result then follows from the Central Limit Theorem. □

Lemma S15. *Under the assumptions of Theorem 6, it holds that*

$$\sum_{j=1}^{M} \frac{\lambda_j}{\delta_j^2} = O\left(\sum_{j=1}^{M} \frac{1}{\lambda_j \delta_j^2}\right),$$

as $n \to \infty$.

Proof of Lemma S15. Since $\lambda_j \to 0$ as $j \to \infty$, there exists $J^* \geq 1$ such that $\lambda_j \geq 1$ for $j \leq J^*$ and $\lambda_j < 1$ whenever $j > J^*$. Note that

$$\begin{aligned}
\sum_{j=1}^{M} \frac{\lambda_j}{\delta_j^2} &= \sum_{j=1}^{M} \frac{1}{\lambda_j \delta_j^2} + \sum_{j=1}^{M} \left(\lambda_j - \frac{1}{\lambda_j}\right) \frac{1}{\delta_j^2} \\
&= \sum_{j=1}^{M} \frac{1}{\lambda_j \delta_j^2} + \sum_{j=1}^{J^*} \left(\lambda_j - \frac{1}{\lambda_j}\right) \frac{1}{\delta_j^2} + \sum_{j=J^*+1}^{M} \left(\lambda_j - \frac{1}{\lambda_j}\right) \frac{1}{\delta_j^2}, \quad \text{(S.156)}
\end{aligned}$$

whence it suffices to show that the third term in (S.156) diverges to $-\infty$ as $n \to \infty$. For this,

$$\sum_{j=J^*+1}^{M} \left(\lambda_j - \frac{1}{\lambda_j}\right) \frac{1}{\delta_j^2} \leq \lambda_{J^*+1}^2 \sum_{j=J^*+1}^{M} \frac{1}{\lambda_j \delta_j^2} - \sum_{j=J^*+1}^{M} \frac{1}{\lambda_j \delta_j^2} = \sum_{j=J^*+1}^{M} \frac{1}{\lambda_j \delta_j^2} \left(\lambda_{J^*+1}^2 - 1\right).$$

The result follows from the fact that $\lambda_{J^*+1}^2 - 1 < 0$ and since $\sum_{j=1}^{M} \lambda_j^{-1/2}\delta_j^{-1} \to \infty$ as $n \to \infty$ implies $\sum_{j=J^*+1}^{M} \lambda_j^{-1}\delta_j^{-2} \to \infty$ as $n \to \infty$. □

Consider the Brownian motion as an example of a Gaussian process for which $\lambda_m = 4/(\pi^2(2m-1)^2)$ and $\phi_m(t) = \sqrt{2}\sin((2m-1)\pi t/2)$ (Hsing and Eubank, 2015). Adopting the optimal bandwidth choices as discussed in Section 4 leads to $c_n \asymp (\log(n)/n)^{1/3}$.

Lemma S16. *Let $\rho \in (1/3, 1)$. For the Brownian motion, if $M = M(n)$ satisfies*

$$M(n) \asymp \left(\frac{\log(n)}{n}\right)^{(\rho-1)/15}, \tag{S.157}$$

then condition (B3) holds and

$$\tau_M \asymp \left(\frac{\log(n)}{n}\right)^{(\rho-1)/5}, \tag{S.158}$$

$$\upsilon_M \asymp \left(\frac{\log(n)}{n}\right)^{4(\rho-1)/15}. \tag{S.159}$$

Moreover, if $\sigma_m^2 \leq Cm^{-(8+\delta)}$ for some constant $C > 0$ and $\delta > 0$, then (B2) is satisfied, $\Theta_M = O\left(M^{-(1+\delta/2)}\right)$ and the rate α_n in Theorem 5 satisfies the following conditions: If $\rho \leq (5+\delta)/(15+\delta)$, then $\alpha_n = O((\log(n)/n)^{(13\rho-3)/30})$ while if $\rho > (5+\delta)/(15+\delta)$ it holds that $\alpha_n = O((\log(n)/n)^{(1-\rho)(1+\delta/2)/15})$. The optimal rate is achieved when $\rho = (5+\delta)/(15+\delta)$ and leads to $\alpha_n = O((\log(n)/n)^q)$, where $q = ((2+\delta)/(15+\delta))/3$.

Proof of Lemma S16. For any $m \geq 1$

$$\lambda_m - \lambda_{m+1} = \frac{32}{\pi^2}\frac{m}{(2m-1)^2(2m+1)^2},$$

which is decreasing as $1 \le m \to \infty$ and thus the eigengaps are given by

$$\delta_m = \frac{32}{\pi^2}\frac{m}{(2m-1)^2(2m+1)^2}, \quad m \ge 1.$$

Since the harmonic sum $H(M) = \sum_{m=1}^{M} 1/m$ satisfies $H(M) \le 1 + \log(M)$ and $M = M(n) \to \infty$ as $n \to \infty$, we obtain

$$\sum_{m=1}^{M} \frac{1}{\sqrt{\lambda_m}\delta_m} = \frac{\pi^3}{64}\sum_{m=1}^{M}\frac{(2m-1)^3(2m+1)^2}{m} \asymp M(n)^5.$$

If $M = M(n)$ satisfies (S.157), then $\sum_{m=1}^{M} \lambda_m^{-1/2}\delta_m^{-1} \asymp c_n^{\rho-1}$ and thus condition (B3) is satisfied. A simple calculation leads to

$$\tau_M = \sum_{m=1}^{M}\frac{1}{\lambda_m} = \sum_{m=1}^{M}\frac{\pi^2(2m-1)^2}{4} \asymp M(n)^3,$$

and

$$\upsilon_M = \sum_{m=1}^{M}\frac{1}{\delta_m} = \frac{\pi^2}{32}\sum_{m=1}^{M}\frac{(2m-1)^2(2m+1)^2}{m} \asymp M(n)^4.$$

The results in (S.158) and (S.159) then follow. If $\sigma_m^2 \le Cm^{-(8+\delta)}$ for some $C, \delta > 0$, then $\sum_{m=1}^{\infty} \sigma_m^2/\lambda_m^2 \le O(1)\sum_{m=1}^{\infty} m^{-(4+\delta)} < \infty$ and condition (B2) is satisfied. From the orthonormality of the ϕ_m

$$\begin{aligned}
\Theta_M = \Big\| \sum_{m \ge M+1} \frac{\sigma_m}{\lambda_m}\phi_m \Big\|_{L^2} \le \sum_{m \ge M+1}\frac{|\sigma_m|}{\lambda_m} &\le O(1)\sum_{m \ge M+1} m^{-(2+\delta/2)} \\
&\le O(1)\int_M^{\infty} s^{-(2+\delta/2)}ds \\
&= O\left(\frac{1}{M^{1+\delta/2}}\right),
\end{aligned}$$

which implies $\Theta_M = (\log(n)/n)^{(1-\rho)(1+\delta/2)/15}$. Also note that $c_n\upsilon_M \asymp (\log(n)/n)^{(1+4\rho)/15}$ and $c_n^{\rho}\tau_M^{1/2} \asymp (\log(n)/n)^{(13\rho-3)/30}$. This implies

$$\alpha_n = c_n\upsilon_M + c_n^{\rho}\tau_M^{1/2} + \Theta_M \le O((\log(n)/n)^{(13\rho-3)/30} + (\log(n)/n)^{(1-\rho)(1+\delta/2)/15}).$$

Thus, if $\rho \leq (5+\delta)/(15+\delta)$, then $\alpha_n = O((\log(n)/n)^{(13\rho-3)/30})$. Similarly, if $\rho > (5+\delta)/(15+\delta)$, then $\alpha_n = O((\log(n)/n)^{(1-\rho)(1+\delta/2)/15})$. The optimal rate is achieved when $\rho = (5+\delta)/(15+\delta) \in (1/3, 1)$ and leads to $\alpha_n = O((\log(n)/n)^q)$, where $q = ((2+\delta)/(15+\delta))/3$. □

References

Amari, S.-i. and Matsuda, T. (2021). Wasserstein statistics in one-dimensional location scale models. *Annals of the Institute of Statistical Mathematics*.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. New York: Springer-Verlag.

Dai, X., Müller, H.-G. and Tao, W. (2018). Derivative principal component analysis for representing the time dynamics of longitudinal and functional data. *Statistica Sinica* **28**, 1583–1609.

Facer, M. R. and Müller, H.-G. (2003). Nonparametric estimation of the location of a maximum in a response surface. *Journal of Multivariate Analysis* **87**, 191–217.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.

Zhang, X. and Wang, J.-L. (2016). From sparse to dense functional data and beyond. *Annals of Statistics* **44**, 2281–2321.